

IEMA

**Grupo de Innovación en la
Evaluación para la Mejora del
Aprendizaje Activo**

**La evaluación compartida:
investigación multidisciplinar**

Editoras

***Frances Watts
Amparo García-Carbonell***

© Edición: Frances Watts y Amparo García Carbonell

© Textos: los autores

© Portada: Jim Rowe <http://www3.sympatico.ca/jim.rowe2/>

Edita: EDITORIAL DE LA UPV
Camino de Vera, s/n
46071 VALENCIA
Tel. 96 387 70 12
Fax 96 387 79 12

Imprime: REPROVAL, S.L.
Tel. 96 369 22 72

Depósito Legal: V-3000-2006
ISBN: 84-7721-633-9

Índice

Autores	
Prólogo	i
Agradecimientos	
Introducción	1
Frances Watts, Amparo García-Carbonell y Juan Llorens	
Alumnos y profesores como evaluadores de presentaciones orales	11
Juan Marin-García	
Investigación cualitativa y cuantitativa en la coevaluación: correlación profesor-alumnos	31
Begoña Montero-Fleta	
Guía para la coevaluación del trabajo en equipo	47
Fernando González Ladrón de Guevara	
Evaluación, coevaluación y autoevaluación del trabajo en grupo en la lectura de mapas topográficos	69
M ^a Ángeles Andreu-Andrés y Miguel García-Casas	
Evaluación de trabajo en grupo en el área de matemáticas	91
M ^a José Pérez Peñalver	
Relación entre autoevaluación y evaluación por el profesor. Estimación del tiempo de dedicación del estudiante	109
Frances Watts, Amparo García-Carbonell y Nieves Martínez Alzamora	
Investigación sobre la evaluación en un curso universitario de química general	131
Juan Llorens	
Evaluación continua y formativa en un curso universitario de química	165
Miguel Anglés, Mónica Edwards, Rafael Llopis y Juan Llorens	

Autores

M^a Angeles Andreu-Andrés

Departamento de Lingüística Aplicada, grupo de investigación DI-AAL
Universidad Politécnica de Valencia (UPV); maandreu@idm.upv.es

Miguel Anglés

Escuela Técnica Superior del Medio Rural y Enología
Universidad Politécnica de Valencia (UPV); m1angles@gmail.com

Mónica Edwards

Escuela Técnica Superior de Ingeniería del Diseño
Universidad Politécnica de Valencia (UPV); moed@doctor.upv.es

Amparo García-Carbonell

Departamento de Lingüística Aplicada, grupo de investigación DI-AAL
Universidad Politécnica de Valencia (UPV); agarcia@upvnet.upv.es,
<http://www.upv.es/diaal/>

Miguel García-Casas

Instituto de Enseñanza Secundaria La Morería, Mislata (Valencia);
mgcasas5@yahoo.es

Fernando González Ladrón de Guevara

Departamento de Organización de Empresas, grupo de investigación ITIO
Universidad Politécnica de Valencia (UPV); fgonzal@omp.upv.es

Rafael Llopis

Departamento de Ingeniería Textil y Papelera
Universidad Politécnica de Valencia (UPV); rllopis@txp.upv.es

Juan Llorens

Departamento de Química
Universidad Politécnica de Valencia (UPV); juallom2@qim.upv.es

Juan Marín-García

Departamento de Organización de Empresas
Universidad Politécnica de Valencia (UPV); jamarin@omp.upv.es

Nieves Martínez Alzamora

Departamento de Estadística e Investigación Operativa Aplicadas y Calidad
Universidad Politécnica de Valencia (UPV); nalzamor@eio.upv.es

Begoña Montero-Fleta

Departamento Lingüística Aplicada, grupo de investigación DI-AAL
Universidad Politécnica Valencia (UPV); bmontero@idm.upv.es

María José Pérez Peñalver

Departamento de Matemática Aplicada,
grupo de investigación Topología y sus Aplicaciones
Universidad Politécnica de Valencia (UPV); mjperez@mat.upv.es

Frances Watts

Departamento de Lingüística Aplicada, grupo de investigación DI-AAL
Universidad Politécnica de Valencia, (UPV); fwatts@upvnet.upv.es

Prólogo

El propósito más importante de la evaluación no es demostrar sino perfeccionar...

Stufflebeam, D.L. y Shinkfield, A. (1993). *Evaluación sistemática. Guía teórica y práctica*. Barcelona: Paidós/M.E.C.

Las claves del interés y utilidad de este libro las encontraremos rápidamente si nos fijamos tanto en el título de esta publicación como en la denominación del grupo de profesores-investigadores. Los autores son miembros del Grupo de Innovación en la Evaluación para la Mejora del Aprendizaje Activo (IEMA) y el título nos remite a la evaluación compartida, es decir, a los alumnos tomando parte activa en su propio proceso de aprendizaje y evaluación.

IEMA es un grupo de profesores universitarios de distintas disciplinas que ha sabido aunar docencia e investigación, indagando sobre su propia docencia, tal como se nos explica en la introducción. Aunque cada capítulo sea fruto del trabajo de un profesor, o de dos o tres asociados para la misma investigación, un grupo constituido con esta finalidad da un apoyo moral, motivacional y académico a las personas que lo integran que, a su vez, facilita tanto su productividad personal como el que una obra como la que aquí se presenta vea la luz para ilustración de muchos otros. Esta labor es fruto de un trabajo que no debemos pasar por alto por lo que tiene de estímulo y de ejemplo.

Otra característica relevante de esta publicación es que es muy coherente con las innovaciones que nos trae la organización académica del nuevo Espacio Europeo, centrada en el aprendizaje y necesariamente muy abierta a la innovación. En este sentido, éste es un libro que nos llega en un momento muy oportuno porque nos puede ayudar en la necesaria adaptación pedagógica a las nuevas orientaciones.

Destaco también como algo importante que se trata de una obra multidisciplinar. Tendemos a concebir la investigación como una actividad propia de los contenidos de una disciplina académica; así el campo habitual de investigación de un

químico es la química. Aquí, en cambio, lo que vemos es que también la química, como cualquier otra disciplina, puede y debe ser objeto de investigación en cuanto enseñada y aprendida. Esta orientación no es patrimonio de una única especialidad académica y aquí tenemos un buen ejemplo de ello. Cualquier profesor de cualquier ámbito académico puede tener como objeto de su investigación su propia tarea docente y el aprendizaje de sus alumnos.

Los nuevos planteamientos que se desprenden del paradigma europeo nos dicen que en la enseñanza universitaria el acento ya no está en el profesor que enseña sino en el alumno que aprende. Podemos definir la enseñanza como una actividad que proporciona a los alumnos oportunidades para aprender; nuestra tarea como profesores está cumplida no si nosotros hemos enseñado, sino si nuestros alumnos han aprendido. Ciertamente, dar más relieve al trabajo del alumno es más coherente con un papel del profesor más centrado en la ayuda al aprendizaje. Este profesor-facilitador, presente de manera inequívoca en las investigaciones que muestra esta publicación, supone necesariamente innovación en las metodologías didácticas, menos centradas en la lección magistral y más en la experimentación. La experimentación es necesaria porque no estamos ante una ciencia exacta y, además, las situaciones, tipos de alumnos y disciplinas que enseñamos pueden ser muy distintas. Aquí entra por lo tanto, y por derecho propio, la investigación sobre nuestra propia acción docente, como se aprecia en cada capítulo. Investigar sobre lo que hacemos en el aula y sobre lo que aprenden nuestros alumnos es, en definitiva, una evaluación de nuestra tarea docente y de nuestras innovaciones didácticas. Unas innovaciones nos pueden salir bien y otras no tanto; la reflexión sobre los resultados, y en eso desemboca la investigación en el aula, es lo que nos irá dando pautas eficaces de acción para que nuestros alumnos aprendan. La transición de un paradigma heredado, en el que el sujeto importante es el profesor que enseña, al nuevo paradigma en el que el énfasis está en el alumno como sujeto activo de su propio aprendizaje, es un proceso de cambio gradual que requiere experimentación.

Quiero señalar, además, que este investigar sobre el aprendizaje del alumno estimula nuestra creatividad al tener que especificar los objetivos de nuestra investigación y hace que nos planteemos estrategias docentes y de evaluación que quizás no nos habríamos planteado de seguir con un modelo más tradicional centrado en lo que hace el profesor más que en lo que hacen y aprenden los alumnos. En este sentido este libro es una buena fuente de sugerencias didácticas, no sólo de investigación, sobre qué podemos promover para un aprendizaje más activo.

Me parece importante añadir que el hecho de investigar sobre todos los aspectos del proceso de enseñanza y aprendizaje supone plantearnos unos objetivos concretos, revisar y estudiar los resultados de experiencias análogas para situar en el contexto apropiado la propia investigación (de esto tenemos un buen ejemplo en la rica

bibliografía de cada capítulo), preparar unos instrumentos o modos de obtener datos coherentes con los objetivos, llevar a cabo los análisis apropiados y reflexionar sobre los resultados. En conjunto, toda esta actividad investigadora puede ser más enriquecedora para el profesor-investigador que muchos seminarios o cursos de formación.

Los objetivos y temas investigados en los distintos capítulos son los que podíamos esperar de un grupo que se autodenomina de innovación en la evaluación para la mejora del aprendizaje; el énfasis está en la evaluación y en el trabajo en grupo. De esta forma conectamos con las exigencias del nuevo paradigma: se trata de un aprendizaje activo, con actividades grupales y de una evaluación colaborativa de los alumnos y del profesor. El que la evaluación sea el objetivo primordial de cada investigación es importante porque es la evaluación la que en último término define los objetivos de aprendizaje y, en definitiva, el cómo estudia el alumno y cómo se forma depende de la evaluación esperada. En los casos presentados la evaluación no es sólo esperada, sino que los mismos alumnos participan en ella, convirtiéndose así la misma evaluación en una nueva experiencia de aprendizaje activo, no ya de conocimientos sino también de habilidades sociales. Se advierte un énfasis en la denominada evaluación formativa, orientada a corregir errores a tiempo, en definitiva orientada a aprender. Solemos asociar la evaluación a exámenes y calificaciones, a verificar lo aprendido; esta concepción de la evaluación hace ya tiempo que está sujeta a revisión. Ahora se habla además de evaluar para ayudar a aprender y las investigaciones aquí presentadas son un buen ejemplo.

Las metodologías que implican una mayor actividad e intervención personal y grupal de los alumnos nos llevan a nuevos enfoques en la evaluación, de manera que se correspondan con los objetivos de aprendizaje. Sobre este aprendizaje colaborativo en la universidad ya va habiendo una abundante literatura experimental, enriquecida ahora por los autores de esta publicación.

Lo que vamos a encontrar en estas páginas tiene que ver con la evaluación de las presentaciones orales y del trabajo en grupo, con la autoevaluación de los alumnos; hay interesantes comparaciones entre cómo se autoevalúan los alumnos y cómo les evalúan sus profesores. Todo lleva a una reflexión y a unas conclusiones que ciertamente van a incidir en una mayor calidad en el aprendizaje de los alumnos. Veremos además en cada capítulo un excelente marco conceptual, una revisión de estudios semejantes para situar cada investigación en su propio contexto, una bibliografía muy útil, guías de evaluación de las diversas experiencias y modelos de análisis. Son investigaciones que sirven de ejemplo y sugieren lo que se puede hacer, y a las que esperamos que sigan muchas otras.

La evaluación compartida: investigación multidisciplinar

Se trata, en definitiva, de una obra muy de actualidad porque se enmarca en las orientaciones sobre la enseñanza y el aprendizaje derivadas del proceso de convergencia europeo y, además, sugiere vías concretas para centrar la enseñanza en la actividad y en el aprendizaje del alumno. Tanto el ejemplo de los autores, profesores-investigadores, como las investigaciones multidisciplinarias presentadas nos señalan un camino muy acorde con la nueva realidad.

Pedro Morales

Universidad Pontificia Comillas, Madrid

Agradecimientos

La investigación centrada en un trabajo de campo orientado a la mejora de la evaluación ofrece la oportunidad de indagar en el proceso y en el producto. Por un lado, el estudio del proceso investiga el proceso cognitivo que genera la evaluación en sí misma, pero también analiza el enfoque didáctico. Por otro, la valoración del producto responde no sólo a la cuantificación de resultados, sino que consensúa criterios, analiza y contrasta resultados y permite la coevaluación. Este planteamiento es el que fundamentalmente ha guiado los diferentes estudios del Grupo de Innovación en la Evaluación para la Mejora del Aprendizaje Activo (IEMA) sobre la evaluación del trabajo en grupo que recoge este volumen.

Los miembros del Grupo IEMA, promovido por el Instituto de Ciencias de la Educación de la Universidad Politécnica de Valencia, agradecen su orientación y apoyo técnico, en especial a Amparo Fernández March, por su empeño y empuje. De igual modo, agradecen las facilidades brindadas por el Vicerrectorado de Estudios y Convergencia Europea, en el marco del Plan de Acciones para la Convergencia Europea (PACE), que impulsa la Universidad Politécnica de Valencia. El grupo IEMA quiere también expresar el debido reconocimiento a Carlos Torrecillas López, becario, por su inestimable trabajo y dedicación.

Una mención especial requiere Pedro Morales, de la Universidad Pontificia Comillas, por compartir su maestría en el campo de la evaluación.

Introducción

Frances Watts
Amparo García-Carbonell
Juan Llorens

Los continuos cambios en el paradigma educativo requieren la implantación de nuevas orientaciones en la enseñanza-aprendizaje, exigen aunar los esfuerzos de todos los implicados y hace imprescindible la búsqueda en común de una respuesta eficaz a la incertidumbre que se pueda generar. El uso de metodologías activas en el aula para estimular y mejorar el aprendizaje conlleva nuevos enfoques en la evaluación que se correspondan con los objetivos de aprendizaje. Estos enfoques requieren definir unas estrategias de evaluación de forma que, además de servir como impulso para la mejora de la adquisición del conocimiento, constituyan un proceso de aprendizaje en sí mismas. El grupo IEMA recopila en este volumen diferentes trabajos que exploran formas distintas de valorar y calificar el rendimiento en distintas asignaturas, en las que la estrategia docente se centra principalmente en el trabajo en grupo y en la evaluación formativa.

Cualquier proceso de transformación socieducativo requiere una reflexión desde la práctica que permita en cada momento reconducir el cambio en función de los objetivos. La permuta educativa que se plantea supone detectar y acotar las deficiencias que puedan surgir en el proceso de formación, introducir los nuevos presupuestos teóricos y realizar un análisis crítico de las estrategias docentes utilizadas, perfilando y asentando las herramientas pedagógicas en las que se concretan tales innovaciones. De igual modo, es necesario establecer algunas conclusiones, siempre abiertas a posibles cambios, para plantear las futuras líneas de trabajo y perfilar estrategias que permitan seguir avanzando. Se trata, en definitiva, de implementar en la práctica educativa una evaluación de calidad, en este caso de un servicio público de Educación Superior.

El enfoque en la evaluación que se plantea tiene importantes implicaciones, ya que sitúa la actividad evaluadora en un contexto mucho más amplio. Realidades como la organización curricular horizontal y vertical de los centros o la evaluación de las tareas del docente, entre otros, no son ajenos al diseño de la evaluación en los diferentes contextos específicos de enseñanza-aprendizaje; se trata de delimitar claramente responsabilidades y competencias. De este modo, la evaluación en sí misma se contempla como un instrumento de aprendizaje, donde un conjunto de actividades son capaces de proporcionar situaciones de realimentación eficaz -evaluación formativa- tanto para los alumnos como para los profesores, al tiempo que inciden en los contenidos, la organización en el aula o los materiales didácticos, entre otros.

Desde el punto de vista didáctico, el impacto de la evaluación afecta a todas las facetas del aprendizaje, desde la motivación e intereses de los alumnos a la actividad desarrollada en el aula (Crooks, 1988) se consideran fundamentales y constituyen la premisa base de los trabajos que aquí se presentan. En definitiva, las innovaciones introducidas en la evaluación inciden directamente en los objetivos y la metodología de enseñanza-aprendizaje. Un hecho claramente constatable desde la práctica docente es que el contenido de las actividades de evaluación y el conocimiento claro de los criterios de valoración de las mismas constituyen un mensaje que ha de ser claro y directo para el alumno, mediante el cual puede entender los objetivos de aprendizaje: qué se considera importante, qué tipo de destrezas y habilidades hay que perseguir, etc. No es posible plantear la evaluación fuera del marco del proceso de formación o al margen de un análisis crítico acerca de sus fines. En este sentido, un reto esencial en cualquier propuesta educativa innovadora es conjugar adecuadamente reflexión teórica y aplicación práctica.

El trabajo en grupo

En esencia, el trabajo en grupo implica el adquirir contenidos y producir resultados en común. Autores como Adams y Hamm (1998) afirman que la labor cooperativa es un método de indagación científica que permite a los alumnos observar fenómenos y comprender las realidades del universo. Al compartir ideas y recursos los problemas y las preguntas se convierten en herramientas didácticas para el descubrimiento. La investigación sobre la efectividad del trabajo en grupo en comparación con esquemas de trabajo individual, según el meta-análisis de Johnson, Johnson y Stanne (2000), demuestra que el trabajo cooperativo renta resultados cognitivos más altos. La colaboración con otros alumnos propicia un aprendizaje más proactivo e independiente, al tiempo que permite abarcar de forma colectiva un grado mayor de contenidos que lo que ofrece una visión individualizada (Higher Education Academy, 2004).

El trabajo cooperativo procura un aprendizaje intenso y significativo, ya que permite que el alumno adquiera un conjunto de competencias interpersonales, instrumentales y sistémicas. Facilita las destrezas interpersonales para la colaboración y el trabajo en equipo, tales como la negociación, la resolución de conflictos y problemas, la toma de decisiones y la revisión constructiva que permite avanzar en los procesos; estas destrezas son un valor añadido en la vida académica, profesional y personal. El trabajo en equipo enriquece especialmente las tareas complejas al aportar diferentes perspectivas y habilidades de los componentes del grupo que se complementan entre sí (Crooks, 1988). A todo ello cabe añadir que la especialización y la educación globalizada es ya una realidad en la que poco hay que no requiera del concurso interdisciplinar.

El trabajo en grupo es una plataforma de encuentro donde los alumnos se conocen y forman relaciones de trabajo que les aportan beneficios más allá de las tareas asignadas. Se ejercitan las estrategias cooperativas propuestas por Axelrod (1984) con las que todos ganan frente a estrategias donde unos ganan y otros pierden o todos pierden; se genera incluso un “capital social” de la clase (Putman, 1993). La clave es compartir (Tinto, 1998), de este modo se forma una comunidad de aprendizaje en la que se comparte la inquietud y el conocimiento sobre un tema, los valores sociales e intelectuales y la responsabilidad de llevar a cabo las tareas planteadas.

La mayoría de los alumnos considera que el trabajo en grupo es más agradable, siendo ésta una de las razones por lo que se incrementa la motivación intrínseca en ellos (Crooks, 1988). Investigaciones sobre la efectividad educativa demuestran que el sentido de comunidad de aprendizaje previene el abandono y cumple la función de incrementar el flujo de información entre alumnos, la disponibilidad de apoyo, el compromiso con los objetivos del grupo, la cooperación entre sus integrantes y la satisfacción general con los esfuerzos del grupo (Rovai, 2002).

El estudio del comportamiento de las variables que intervienen en la evaluación del trabajo en equipo se convierte pues en una labor esencial. Si los procesos y los resultados no se miden y valoran adecuadamente, la eficacia del método no está garantizada.

Variables de la evaluación

Las ventajas que ofrece el trabajo cooperativo invitan a los docentes a plantearlo como estrategia básica de enseñanza-aprendizaje. El reto es superar los inconvenientes de evaluar el progreso en el conocimiento y las distintas habilidades adquiridas. Para ello, un primer paso es conocer bien los aspectos que deben considerarse para una evaluación más fiable.

Es importante establecer adecuadamente una serie de parámetros antes de emprender una metodología basada en el trabajo cooperativo. El tamaño del grupo ha de ir en función de características propias de la disciplina y del tipo de tarea. En cualquier caso, si los grupos suponen más de tres o cuatro personas se tiende, por lo general, a perder la entidad de equipo y a trabajar en subgrupos. Una segunda cuestión es la duración de la tarea, que puede ocupar una sesión de clase, un día, una semana, un semestre o todo un curso. Si la duración es prolongada, exige marcar los objetivos intermedios, pautas de revisión y acuerdo con los alumnos. La tercera cuestión es la formación de los grupos de trabajo; si los forma el profesor ha de explicitar los criterios para la agrupación, por ejemplo, una agrupación basada en el nivel de conocimientos similares. La cuarta cuestión es la especificación de las tareas a realizar, bien por el profesor, por el grupo de trabajo o por consenso de ambas partes. Una última cuestión es determinar qué producto se persigue: un informe de la tarea o proyecto, una presentación oral, una exposición, un objeto, un portafolio, o quizás alguna combinación de éstos (adaptado de OCSLD, 2002).

En la evaluación del trabajo en grupo se plantean las mismas cuestiones de *qué, quién, cómo, cuándo y para qué* (Rodríguez Neira et al., 1995) que plantean otros tipos de evaluación. Las preguntas planteadas de forma precisa serían: ¿se evalúa el proceso o el producto del trabajo en grupo, o una combinación de ambos?; ¿qué tipo de evidencia o prueba se evalúa en el producto, un informe colectivo, informes individuales, una práctica individual o del grupo, una simulación?; ¿quién evalúa qué? y ¿cuándo tiene lugar la evaluación?

Desde el punto de vista del proceso, la evaluación del trabajo en grupo ha de servir como una constatación continua de la realización de las tareas. Stufflebeam y Shinkfield (1993) apuntan que otro objetivo es comprobar hasta qué punto los participantes aceptan y son capaces de asumir sus responsabilidades y desempeñar sus tareas. Esta valoración se puede llevar a cabo por los mismos componentes del grupo de trabajo, bien evaluándose los unos a los otros (evaluación entre pares o *peer assessment*) o bien cada uno a sí mismo (autoevaluación o *self assessment*). Este enfoque puede aportar un efecto formativo considerable por el incentivo que supone gestionar e incidir directamente en el progreso del grupo. Esta estrategia puede contribuir a corregir algunas anomalías y carencias que a menudo se producen en el trabajo colectivo. Por ejemplo, un alumno puede dominar a otras personas del grupo en detrimento de los más tímidos e inseguros, o puede agazaparse en el grupo sin trabajar y oculto a los ojos del profesor. La detección temprana o antes de avanzar demasiado en las tareas, mediante encuesta o puesta en común de los componentes del grupo y con la ayuda del profesor, puede paliar deficiencias que pueden resquebrajar el grupo y conducir al fracaso.

Por otra parte, si la evaluación del proceso constituye parte de la calificación final individual, a los alumnos les preocupa que la calificación refleje el nivel de

contribución personal. Existe la posibilidad de que se otorgue una calificación a cada miembro del grupo y que todos los componentes reciban la media de las calificaciones individuales. El profesor se puede apoyar en los diarios de actividad llevados por los grupos, en las actas de las reuniones de grupo y en la observación directa del proceso. Es fundamental que los criterios de evaluación estén predeterminados, bien por el profesor o consensuados con los alumnos. Tanto la claridad de criterios como el control del trabajo del grupo ayudan a los alumnos a centrarse en el proceso, que en sí mismo ya constituye una experiencia de aprendizaje (James, McInnis y Devlin, 2002). Sin embargo, la información obtenida con la evaluación del proceso puede tener tintes subjetivos y, por tanto, ofrecer problemas de fiabilidad, además, supone una carga adicional en el tiempo de dedicación del profesor. Una consideración importante es que los premios, becas o cualquier otro beneficio que el estudiante puede alcanzar se conceden sobre la base de las calificaciones de los aspirantes, por lo que se debe meditar antes de otorgar una calificación uniforme para todos los componentes del grupo en la calificación final de una asignatura. Por esta última razón, Isaacs (2002) aboga por una calificación individual; afirma que la calificación individual premia a los alumnos destacados y penaliza la actuación de los sujetos-parásitos.

Si se pretende evaluar el producto, Stufflebeam y Shinkfield (1993) opinan que se debe aplicar diferentes técnicas de valoración, ya que no existe una fórmula única válida para todos los casos. Lo deseable es proporcionar el mayor número de comprobaciones de los resultados de aprendizaje y conseguir una visión amplia de los logros. Por ejemplo, puede evaluarse el producto del grupo y otorgarse la misma calificación a cada componente. Esto premia la efectividad de la colaboración entre los componentes pero a los alumnos con mayor dedicación puede parecerles injusto que a todos se les califique igual y con ello premiar a los de menos dedicación. De igual modo, se le puede otorgar una única calificación al producto en sí mismo, que el grupo puede distribuir individualmente por medio de puntos según criterios consensuados previamente. Si la asignación individual de calificaciones se justifica y se basa enteramente en los criterios, se reduce el peligro de que se premie a los alumnos con habilidad para la negociación. Una tercera manera de dividir la valoración del trabajo en grupo es que cada componente pida que se le asigne una parte definida del trabajo y que se le evalúe según su parte. Esta manera de proceder en la evaluación puede parecer más justa pero no alienta la colaboración. Los procedimientos mencionados pueden combinarse de varios modos. En el caso de que los alumnos reciban la misma calificación por el producto del trabajo en grupo, a la vez pueden recibir una calificación por su contribución individual, la cual puede ser evaluada por el profesor al observar el grupo durante el proceso y por la reflexión de cada componente sobre el proyecto o tarea, sobre el proceso y sobre su dedicación y lo que han aprendido.

Es factible también que los alumnos reciban una calificación grupal pero que se asigna en las diferentes fases o actividades del trabajo. El grupo entonces acuerda quienes han de recibir más o menos puntos en cada etapa o tarea. Para mantener el promedio los alumnos comparan y ponderan sus contribuciones. Una breve entrevista con cada persona del grupo también dará al profesor idea de la naturaleza e intensidad de la contribución de cada alumno respecto del grupo. El profesor podría ajustar la calificación del grupo, por ejemplo, hasta el 10% en base a la entrevista.

El producto puede ser un informe individual escrito defendido por el alumno en una presentación oral ante el profesor o ante el profesor y los demás alumnos de la clase. Los alumnos trabajan en grupo pero cada uno entrega su propio informe. Esta estrategia de evaluación aporta el beneficio de compartir información y puntos de vista pero los alumnos han de llegar a conclusiones propias.

Cuando se trata de trabajo individual, cabe mencionar el peligro de plagio, entendido éste como la utilización del trabajo de otros sin el reconocimiento de la fuente. El plagio puede ocurrir por varias razones, con la intención de engañar al lector, por descuido o por simple desconocimiento de las convenciones pertinentes. Según Isaacs (2002), es más fácil y preferible hacer que el plagio sea improbable que tener que tratar el hecho después de producirse. Para ello es conveniente advertir a los alumnos de lo que es plagio y que es equiparable a copiar en un examen. Hay que enseñar la forma correcta de reconocer y referenciar el trabajo ajeno, así como los límites aceptables de colaboración entre compañeros en la producción individual.

Por último, como complemento, o incluso en sustitución de los procedimientos mencionados, para calificar el producto se puede confeccionar un examen en el que los alumnos describan y analicen el proyecto de trabajo y su contribución. No hay que olvidar que los exámenes, si se clasifican según el tipo de preguntas, se dividen en pruebas objetivas, pruebas libres y pruebas diagnósticas. La diferencia entre los tres tipos estriba en el grado de libertad del alumno para contestar a las preguntas, siendo en las pruebas objetivas más restringida, mientras que el grado de libertad en la respuesta en las de diagnóstico (observación y encuesta) es superior (Rodríguez Neira et al., 1995). Quizás sea el término medio la mejor vía de examinar el trabajo en grupo. La respuesta abierta, oral o escrita, una prueba práctica real o una simulación, o bien pruebas mixtas, objetivas y libres, pueden sustituir o completar la evaluación del producto del trabajo en grupo. En todo caso, los criterios de corrección deben ser explícitos y públicos desde el principio.

Al examinar los procedimientos para la evaluación del trabajo en grupo, Rising (2004) encuentra veintiún métodos, en muchos casos combinaciones, que contemplan las posibilidades de evaluar el producto y el proceso. También se analiza

que la evaluación derive del profesor y del alumno(s) y la forma de otorgar la calificación al individuo o al grupo. La mayoría de los métodos examinados evalúan el producto, se llevan a cabo por el profesor y conducen a calificaciones que se diferencian entre los individuos.

En resumen, la estructura de la evaluación ha de centrarse en cuatro factores que inciden en las variables que intervienen en ella:

1. Producto o proceso. ¿Se va a evaluar el producto del trabajo en grupo, el proceso del trabajo en grupo o ambos y en qué proporción?
2. Criterios. ¿Qué criterios se van a utilizar para evaluar las tareas en grupo? ¿Quién determinará estos criterios, el profesor, los alumnos o ambos?
3. Evaluador. ¿Quién va a aplicar los criterios de evaluación y determinar las calificaciones? (el profesor, los alumnos entre pares, autoevaluación o una combinación).
4. Asignación de las calificaciones. ¿Una nota del grupo compartida, el promedio del grupo, una calificación individual o una combinación de estos indicadores? (adaptado de James et al., 2002).

Acertar en la estrategia de evaluación influye decisivamente en el éxito del trabajo en grupo. Sea cual fuere el tipo de evaluación que se aplique, es primordial que los alumnos tengan claro desde el principio el tipo de evaluación, sus detalles y su racionalidad. El trabajo en grupo es aconsejable pedagógicamente y profesionalmente, pero requiere una planificación adecuada, clara, guiada y una evaluación fiable.

Este volumen

La colección de trabajos que alberga esta obra es el resultado de diferentes estudios que pueden colaborar en la definición, control y evaluación de tareas basadas en el trabajo cooperativo, en definitiva, en el trabajo en grupo. Se trata de experiencias multidisciplinares de evaluación con un denominador común, el trabajo en equipo como estrategia docente o de evaluación.

Los diferentes estudios ofrecen un análisis de la evaluación tanto del proceso como del producto del trabajo en grupo. En la evaluación del proceso, la mayoría de las aportaciones contemplan la implicación de los alumnos en el consenso de los criterios de observación para aplicarlos en la evaluación entre pares. En la evaluación del producto, existe una variedad de criterios de valoración desarrollados sólo por el profesor o conjuntamente con los alumnos. También hay diversidad entre los casos que aplican los criterios en la evaluación del producto: el profesor,

el propio alumno a sí mismo o a sus iguales. En definitiva, en la mayoría de los estudios se plantea el trabajo en grupo como estrategia docente y/o de evaluación, por lo que esta colección de trabajos aporta modelos de cómo afrontar, o no, la evaluación del trabajo cooperativo y muestra una valiosa recopilación de criterios, parrillas de evaluación y referencias bibliográficas. Palabras clave como trabajo cooperativo, trabajo en grupo, evaluación formativa y sumativa, coevaluación, evaluación entre pares, autoevaluación, evaluación del proceso y del producto, validez y fiabilidad, entre otras, se acomodan en las diferentes aportaciones.

Los profesores-investigadores que integran el grupo IEMA llevan a cabo una labor de estudio e investigación constante sobre la evaluación y los factores que inciden, actúan y se derivan de ella y presentan en este libro experiencias en las que los interesados pueden ahondar si así lo desean.

Bibliografía

- Adams, D. y Hamm, M. (1998). *Collaborative Inquiry in Science, Math and Technology*. Portsmouth, NH: Heinemann.
- Axelrod, R. (1984). *The evolution of cooperation*. Nueva York: Basic Books.
- Crooks, T.J. (1988). The impact of classroom evaluation practice on students. *Review of Educational Research*, Winter 1988. (58) 4, pp. 438-481.
- Higher Education Academy. (2004). *Assessment of group work*. www.heacademy.ac.uk/668.htm [ref. del 5 abril 2005].
- Isaacs, G. (2002). *Assessing group tasks*. Teaching & Educational Development Institute, University of Queensland, Australia. <http://www.tedi.uq.edu.au/teaching/assessment/groupass.html> [ref. del 7 de abril de 2005].
- James, R., McInnis, C. y Devlin, M. (2002). *Assessing learning in Australian universities*. Centre for the Study of Higher Education. University of Melbourne, Australia. <http://www.cshe.unimelb.edu.au/assessinglearning/> [ref. del 6 de abril 2005].
- Johnson, D.W., Johnson, R.T. y Stanne, M.B. (2000). Cooperative learning methods: A meta-analysis. <http://www.co-operation.org/pages/cl-methods.html> [ref. del 30 de enero de 2006].
- Oxford Centre for Staff and Learning Development (OCSLD) (2002). *First words*. Oxford Brookes University. <http://www.brookes.ac.uk/services/ocsd/firstwords/> [ref. del 5 abril de 2005].

- Putman, R. (1993). *Making democracy work*. Princeton: Princeton University Press.
- Rising, B. (2004). Evaluation in the context of collaborative/cooperative learning. En Kriz, W. y Eberle, T. (eds.) *Bridging the gap: Transforming knowledge into action through gaming and simulation*. Munich: SAGSAGA, 326-331.
- Rodríguez Neira, T., Álvarez Pérez, L., Cadrecha Caparrós, M.A., Hernández García, J., Luengo García, M.A., Ordoñez Álvarez, J. y Soler Vázquez, E. (1995). *Evaluación de aprendizajes*. Oviedo: ICE, Universidad de Oviedo.
- Rovai, A.P. (2002). *Building a sense of community at a distance*. International Review of Research in Open and Distance Learning (3) 1. www.irrodl.org/content/v3.1/rovai.html [ref. del 5 abril 2005].
- Stufflebeam, D. y Shinkfield, A. (2ª impresión 1993). *Evaluación sistemática. Guía teórica y práctica*. (Trad. del inglés, *Systematic Evaluation* 1985). Barcelona: Ed. Paidós y MEC.
- Tinto, V. (1998). Learning communities: Building gateways to student success. *National Learning and Teaching Forum*, (7) 4. <http://www.ntfl.com/html/lib/suppmat/74tinto.htm> [ref. del 10 abril 2005].

Alumnos y profesores como evaluadores de presentaciones orales

Juan Antonio Marín García

Este estudio analiza los resultados de incorporar a los estudiantes en el proceso de evaluación y las conclusiones a las que se ha llegado tras la experiencia llevada a cabo durante el curso 2004-05. El trabajo de campo se centra en comprobar el grado de acuerdo de las calificaciones de los alumnos en relación a las del profesor. De igual modo, se comprueba la relación de las calificaciones del profesor de la asignatura con las asignadas por cuatro jueces externos. Los resultados obtenidos parecen indicar que los alumnos pueden evaluar adecuadamente las presentaciones orales de sus compañeros, puesto que en sus puntuaciones no se aprecian diferencias significativas con las que propone el profesor de la asignatura.

Palabras clave: *presentaciones orales, acuerdo entre jueces, evaluación entre pares.*

1. Introducción

La educación universitaria se enfrenta en la actualidad a diferentes retos. Entre ellos, el papel que puede desempeñar la evaluación de los aprendizajes de los alumnos y la importancia del desarrollo de determinadas competencias profesionales (Dochy, Segers y Sluijsmans, 1999; MacLellan, 2001; Orsmond, Merry y Reiling, 1996). Una forma de aunar estos dos aspectos es fomentar la participación de los alumnos en los procesos de evaluación de su propio aprendizaje o el de sus compañeros. Sin embargo, la docencia tradicional no suele contemplar esta posibilidad y la evaluación recae fundamentalmente en los profesores, cuyas calificaciones se consideran válidas e indiscutibles (Orsmond et al., 1996). No obstante, cada vez es más habitual que se reconozca la importancia de formar a los alumnos universitarios

para que desarrollen las habilidades necesarias para reflexionar críticamente sobre los resultados de su trabajo y el proceso seguido para completarlo (Kwan y Leung, 1996). No sólo eso, sino que además sean capaces de evaluar también el trabajo de sus compañeros. Esta habilidad les será muy útil en su futuro entorno profesional, sobre todo cuando desempeñen tareas en equipo (MacPherson, 1999).

Las investigaciones publicadas en revistas científicas que tratan la evaluación de los compañeros suele tener como objeto la puntuación de ensayos, trabajos de grupo o presentaciones orales (Dochy et al., 1999; Magin, 2001b) y, en algunos casos, posters (Orsmond et al., 2000). Como se han comparado actividades muy diferentes, no es de extrañar que presenten resultados divergentes.

El número de investigaciones publicadas hasta la fecha es limitado, por lo que sería recomendable ampliar las experiencias relacionadas con la evaluación de presentaciones orales (MacAlpine, 1999) en diferentes disciplinas (Gatfield, 1999; Kwan y Leung, 1996). Parece necesario incorporar nuevas formas de analizar los datos, que sean más rigurosas y que permitan integrar las conclusiones de diferentes investigaciones (Kwan y Leung, 1996; Magin, 2001a).

En la presente investigación, se ha analizado las formas de incorporar a los estudiantes en el proceso de evaluación y las conclusiones a las que ha llegado la investigación previa. Sólo se han tenido en cuenta los estudios relacionados con la evaluación de presentaciones orales, actividad en la que se centra el trabajo de campo. También se resumen las recomendaciones que se han propuesto para mejorar el proceso de participación de los alumnos en la evaluación. La revisión de la literatura concluye poniendo de manifiesto qué análisis estadísticos se han utilizado para comprobar la fiabilidad de las puntuaciones realizadas por alumnos, los resultados obtenidos en otras investigaciones y las limitaciones de los mismos. El trabajo de campo llevado a cabo se ha centrado en comprobar el grado de acuerdo de las puntuaciones de los alumnos, comparándolas con las del profesor de la asignatura; también se ha comprobado si las puntuaciones del profesor de la asignatura coinciden con las de cuatro jueces más.

1.1. El alumno como evaluador

Existen tres formas de incorporar a los estudiantes en el proceso de evaluación de su aprendizaje (Dochy et al., 1999):

- Autoevaluación (*self-assessment* o SA): consiste en dar a los estudiantes la posibilidad de valorar su propio aprendizaje, en especial, el producto o resultado de su aprendizaje.

- La evaluación entre pares (*peer assessment* o PA): consiste en un proceso a través del cual un grupo de personas puntúa a sus iguales.
- La evaluación participativa (*collaborative assessment, participative assessment, co-assessment* o CA): consiste en dar la oportunidad para que los alumnos se evalúen a sí mismos o entre sí, pero permitiendo al mismo tiempo que el profesor pueda intervenir y controlar la calificación final.

Las tres formas planteadas (SA, PA, CA) se vienen empleando en la educación universitaria desde hace años y han sido objeto de frecuentes investigaciones desde los años 80 (Gatfield, 1999; Searby y Ewers, 1997). Los principales temas a los que se han referido estas investigaciones tienen que ver con el uso de estos procedimientos como evaluación formativa o como evaluación sumativa; el grado de acuerdo de las puntuaciones comparadas profesor-alumno; efectos que produce en los alumnos su participación en la evaluación de su aprendizaje; qué instrumentos de medición se pueden utilizar y cómo se pueden mejorar los procesos de evaluación (Dochy et al., 1999; Gatfield, 1999).

En realidad, los objetivos de las diferentes investigaciones están íntimamente relacionados. Si estas formas de evaluación influyen positivamente en el aprendizaje de los alumnos, está justificado su uso como evaluación formativa. Sin embargo, no sería muy lógico incorporar la evaluación de los alumnos a la evaluación sumativa si el grado de acuerdo con las puntuaciones del profesor o la justicia de la misma no fuese elevada. Por otro lado, los alumnos tienden a prestar más importancia a las cosas que afectan a su calificación final (Magin y Helmore, 2001); es importante identificar qué métodos y bajo qué circunstancias se puede mejorar la calidad de la evaluación de los alumnos para que se pueda incorporar a la evaluación sumativa.

Un repaso por las conclusiones a las que han llegado las investigaciones previas nos muestra un cierto grado de acuerdo en las ventajas que la SA proporciona desde el punto de vista de evaluación formativa, muchas de estas ventajas son compartidas también por la PA. Diferentes autores (Dochy et al., 1999; Kwan y Leung, 1996, entre otros) resumen algunas de estas ventajas: los alumnos tienen más confianza en sus habilidades; mejora la percepción del alumno sobre la calidad de su trabajo; reflexionan más sobre su conducta y sus resultados; mejoran los resultados en los exámenes y la calidad y eficacia del aprendizaje; asumen mayores responsabilidades en el proceso de aprendizaje y la satisfacción aumenta. Además, la SA es muy útil para que los alumnos logren sus objetivos de aprendizaje (Orsmond et al., 1997), hace que sean más críticos, trabajen de manera más estructurada y les anima a pensar más (Kwan y Leung, 1996; Orsmond et al., 2000). Por otra parte, Kwan y Leung (1996) y Reynolds y Trehan (2000) insisten en que desarrollar en los estudiantes la habilidad de evaluar el trabajo propio o el de los

compañeros, es un elemento importante en el proceso de aprendizaje. Es una de las competencias profesionales para las que debemos formar a nuestros universitarios si queremos formar profesionales reflexivos y fomentar el aprendizaje a lo “largo de toda la vida” (Cheng y Warren, 1999; Dochy et al., 1999; MacAlpine, 1999; MacPherson, 1999).

Respecto al grado de acuerdo entre las puntuaciones de los alumnos y profesores, parece que las puntuaciones de los compañeros (PA) son más similares que las de autoevaluación (SA) (Ward, Gruppen y Regehr, 2002). Probablemente esto se deba a que en PA se suelen promediar las puntuaciones de varios alumnos que evalúan a un compañero, antes de compararlas con las del profesor (Magin, 2001b). En todo caso, los principales problemas que se citan acerca del grado de acuerdo afectan a ambas formas de evaluación por parte de los estudiantes. Estos problemas son:

- Se puede pactar la puntuación o puntuar a los amigos o a uno mismo mejor que a los demás (Dochy et al., 1999; Magin, 2001b; Magin y Helmore, 2001).
- Los alumnos no suelen usar todo el rango de la escala disponible y tienen a concentrar sus puntuaciones en el tramo medio (Dochy et al., 1999; MacPherson, 1999).
- Los alumnos no tienen el mismo nivel de referencia que los profesores (Magin y Helmore, 2001).
- Las puntuaciones de los alumnos no se parecen a las de los profesores (Dochy et al., 1999; MacPherson, 1999; Magin y Helmore, 2001).

Debido a los problemas mencionados, la reserva a incorporar las puntuaciones de los alumnos como evaluación sumativa, no es sólo de los profesores. Algunos estudiantes desconfían de la calidad de las puntuaciones de sus compañeros o de ellos mismos y no se sienten cómodos con este sistema (Kwan y Leung, 1996; MacPherson, 1999).

Se puede afirmar que las recomendaciones que hacen los diferentes autores para mejorar el proceso de PA, SA o CA, son congruentes. En este sentido, al poner en marcha estos mecanismos de evaluación es aconsejable tener en cuenta las siguientes observaciones:

- La evaluación es una habilidad que se puede mejorar con la repetición y el entrenamiento (Dochy et al., 1999; Kwan y Leung, 1996; MacPherson, 1999; Magin, 2001a; Searby y Ewers, 1997).

- El alumno ha de contar con tiempo para emitir sus puntuaciones. También es posible que los alumnos necesiten apoyo u orientaciones en algún momento del proceso (Dochy et al., 1999; Magin, 2001a).
- Si se quiere usar SA o PA como herramienta formativa, se debe ayudar a que los estudiantes vean estas tareas como una herramienta que facilita su aprendizaje (Dochy et al., 1999).
- Los evaluadores deben usar los mismos criterios, que estén establecidos de antemano, sean conocidos por los alumnos y estén formulados de una manera clara y sencilla. Si es posible, se recomienda negociarlos con los estudiantes (Cheng y Warren, 1999; Dochy et al., 1999; Kwan y Leung, 1996; MacAlpine, 1999; MacPherson, 1999; Magin, 2001a; Orsmond et al., 2000; Searby y Ewers, 1997; Sullivan y Hall, 1997).
- Aunque las puntuaciones individuales de un juez no sean fiables, cuando se promedian las puntuaciones de varios jueces, la fiabilidad es muy elevada (Magin y Helmore, 2001).
- Usar PA como una parte de la calificación suele ser más fiable que SA y complementa la calificación entregada por los profesores (Dochy et al., 1999; Kwan y Leung, 1996; Magin, 2001a; Searby y Ewers, 1997).

En definitiva, parece que el principal punto de desacuerdo entre las investigaciones es el grado de acuerdo de las puntuaciones de los alumnos. Si se mejorara este aspecto, permitiría reducir algunas de las reservas que hay por parte de los profesores y de los alumnos para adoptar estos sistemas. En este sentido, el objeto del trabajo que aquí se presenta será comprobar el grado de acuerdo que se obtiene en una implantación concreta de PA (evaluación de presentaciones orales), cuando se tienen en cuenta las recomendaciones realizadas por investigaciones previas sobre el tema.

1.2. Grado de acuerdo en las puntuaciones de presentaciones orales

Se han utilizado diferentes medidas para establecer el grado de acuerdo entre las puntuaciones de los alumnos y las de los profesores. La más habitual es usar el coeficiente de correlación entre la media de las puntuaciones de los alumnos con la puntuación del profesor (o la media de las puntuaciones de los profesores, si intervienen varios en la evaluación)¹. Otras medidas utilizadas son calcular el porcentaje de alumnos que dan una puntuación comprendida en un intervalo de

¹ Al-Fallay, 2004; Cheng y Warren, 1999; Falchikov y Goldfinch, 2000; Freeman, 1995; Kwan y Leung, 1996; Langan et al., 2005; MacAlpine, 1999; MacPherson, 1999; Magin y Helmore, 2001.

confianza sobre la puntuación del profesor (normalmente una desviación estándar)², comparar la varianza de las puntuaciones de los alumnos y de los profesores o hacer un T-test para la diferencia de la medias entre las puntuaciones de los alumnos y las de los profesores³. Algunos estudios utilizan una medida asociada al análisis de la varianza (ANOVA) para determinar la fiabilidad entre evaluadores (Magin, 2001a).

El grado de acuerdo entre las puntuaciones promedio de los profesores y las del grupo de alumnos es bastante elevado cuando se evalúan presentaciones orales. Por un lado las correlaciones entre las puntuaciones de los alumnos y las de los profesores son moderadas o altas; los valores se sitúan entre 0,44 y 0,79 en los trabajos de varios autores⁴, mientras que son mayores que 0,80 en otros estudios⁵. Por otro lado, el grado de coincidencia de las puntuaciones de los alumnos y las del profesor es muy elevado 98% (Falchikov, 1995), 95% (Fresan, 1995) y 70% (Kwan y Leung, 1996). También se puede confirmar que no aparecen diferencias significativas entre puntuaciones de alumnos y profesores al aplicar una prueba T de Student (Freeman, 1995; Kwan y Leung, 1996). Sin embargo, la dispersión de las puntuaciones de los alumnos es sensiblemente menor que las de los profesores⁶. En otras palabras, los alumnos tienden a concentrar sus puntuaciones y discriminan menos que los profesores.

No obstante, hay que tener presentes una serie de advertencias realizadas por Ward et al. (2002). En primer lugar, casi todos los estudios consideran que la calificación del profesor es correcta y que la diferencia de puntuaciones con los alumnos es debida a que éstos no cuentan con experiencia puntuando como los profesores. Sin embargo, esta diferencia también se puede deber a que las puntuaciones del profesor no son tan válidas y fiables como cabría pensar (Falchikov y Goldfinch, 2000; Magin y Helmore, 2001; Orsmond et al., 1996). La validez tiene que ver con puntuar lo que se pretende medir y, al mismo tiempo, que lo que se pretende medir sea relevante; que sea representativo de los conocimientos/habilidades objetivo de la asignatura que el alumno ha adquirido. Las sugerencias para superar este problema es usar el promedio de las puntuaciones de varios profesores expertos (Ward et al., 2002).

² Falchikov, 1995; Freeman, 1995; Kwan y Leung, 1996.

³ Cheng y Warren, 1999; Freeman, 1995; Kwan y Leung, 1996; Magin y Helmore, 2001; Ward et al., 2002.

⁴ Cheng y Warren, 1999; Freeman, 1995; Kwan y Leung, 1996; MacPherson, 1999; Magin y Helmore, 2001 y en 3 de los estudios citados por Falchikov y Goldfinch, 2000.

⁵ Al-Fallay, 2004; Langan et al., 2005; MacAlpine, 1999 y en otros 3 estudios citados por Falchikov y Goldfinch, 2000.

⁶ Cheng y Warren, 1999; Freeman, 1995; Kwan y Leung, 1996; Magin y Helmore, 2001.

En segundo lugar, aunque existan una serie de criterios para puntuar, no hay garantía de que todos los evaluadores los interpreten de la misma manera. La mejor forma de evitar este problema es definir los criterios de manera sencilla, incluyendo guías explícitas en la parrilla de calificación (Ward et al., 2002). Otra posibilidad, es tipificar las puntuaciones de los alumnos o de los profesores antes de promediarlas.

Por último, Ward, Gruppen y Regehr (2002) también plantean un posible problema cuando sólo se analizan los datos a nivel de grupo. Es decir, cuando comparamos la media de las puntuaciones de un grupo de evaluadores con la puntuación de un experto. Este tipo de análisis no nos indica el grado de acuerdo de cada alumno individual con la puntuación del profesor y no nos aporta información relevante cuando sólo se dispone de una puntuación otorgada por los alumnos (bien porque se autoevalúan o bien porque cada trabajo es evaluado sólo por un compañero).

En este trabajo, se pretende utilizar las dos primeras recomendaciones de Ward et al. (2002) y se plantea para una investigación posterior la evaluación de los diferentes niveles de análisis.

1.3. Objetivos

Teniendo en cuenta las investigaciones previas revisadas, parece interesante continuar aportando datos para clarificar hasta qué punto se puede confiar en las puntuaciones que proporcionan los alumnos. De este modo, se pretende reflexionar sobre la posibilidad de incorporarlas a la evaluación sumativa de asignaturas universitarias. Las preguntas a las que se pretende responder con esta investigación son las siguientes:

1. ¿Qué grado de acuerdo tienen las puntuaciones de un solo profesor si las comparamos con la de un grupo de profesores?
2. ¿Concuere el promedio de las puntuaciones de varios alumnos evaluando una misma presentación con la puntuación otorgada por el profesor?
3. ¿Cuántos alumnos evaluadores serían necesarios para conseguir una fiabilidad similar a la de un profesor?

2. Método

La respuesta a las preguntas de investigación, se obtiene por los procedimientos siguientes:

- El grado de acuerdo de las puntuaciones de un profesor se mide contrastando las puntuaciones entregadas por el profesor de la asignatura con la media de las puntuaciones de otros cuatro profesores que participan en esta investigación. Para facilitar el proceso de puntuación, se utilizan las presentaciones grabadas en vídeo, tanto para el profesor de la asignatura como para los profesores colaboradores en la investigación. De este modo, se parte de una situación común a todos ellos. Con los datos obtenidos se calcula la correlación y se aplica una prueba de T de Student de diferencia de medias, para comprobar si las puntuaciones se corresponden a una misma población. También se calcula cuántas puntuaciones del profesor de la asignatura se diferencian de las del promedio de los otros cuatro profesores en más de una desviación estándar (Cheng y Warren, 1999; Freeman, 1995; Kwan y Leung, 1996; Magin y Helmore, 2001).
- El grado de acuerdo del promedio de las puntuaciones de los alumnos se intenta probar con un procedimiento análogo. Pero en este caso, el valor de referencia son las puntuaciones otorgadas por el profesor de la asignatura el día de las presentaciones.
- Por último se completarán los análisis con la estimación del número de alumnos o de profesores que deberían evaluar simultáneamente una exposición para que su fiabilidad fuese similar a la de la puntuación de un solo profesor experto (tanto en la materia impartida como en evaluar presentaciones orales) para ello seguiremos el procedimiento sugerido por Magin (2001a) (ver Apéndice A).

Kane y Lawler (1978) sugieren tres procedimientos para puntuar a los compañeros: rango ordenado, nominación y puntuación. En nuestro caso se ha elegido el tercero de ellos, donde cada persona puntúa a sus compañeros de acuerdo con su rendimiento, utilizando una escala de puntuación. Este procedimiento se considera que es el de más fácil aplicación (aunque los otros métodos pueden ser más fiables o permitir un mayor grado de discriminación).

Para fomentar una mayor implicación de los alumnos, se decide que los criterios de puntuación se seleccionen por los estudiantes, tal y como se describe más adelante. La tarea del profesor integra las visiones de los diferentes grupos y genera la versión definitiva de la parrilla, incluyendo criterios para cada aspecto a evaluar. La parrilla contiene 9 aspectos y cada uno de ellos se puntúa entre 0 y 3; la puntuación máxima para una exposición es de 27 puntos.

Los alumnos objeto de estudio están matriculados en uno de los dos grupos de cuarto curso de Ingeniero de Organización Industrial. Ninguno de los alumnos tiene experiencia previa en actividades de evaluación a compañeros. La asignatura *Gestión de Empresas* se imparte durante 15 semanas lectivas, en sesiones de 2,5 horas semanales. La actividad evaluada se lleva a cabo en parejas. Consiste en realizar una entrevista a dos mandos de empresa, comparar sus respuestas con la teoría impartida en el curso y presentar el resultado de su trabajo ante los demás compañeros en clase; exposición que se graba en vídeo. La actividad se expone el último día del curso y no es obligatoria, aunque representa un 10% en la calificación final de la asignatura, el 5% por la puntuación de la exposición (promedio de la puntuación de los compañeros y del profesor) y el 5% por el grado de acuerdo de las puntuaciones puestas por cada alumno al compararla con el promedio de puntuaciones de todos los alumnos. El número de presentaciones asciende a 23 con un total de 44 alumnos participantes.

Un mes antes de la exposición se explica la actividad durante una sesión de clase. Una vez aclaradas las dudas generales se forman dos grupos de 15-20 alumnos, que acuerdan los criterios a utilizar para puntuar la actividad. Se empieza con una “tormenta de ideas con grupo nominal” sobre los aspectos que definen una buena exposición oral. Posteriormente se depuran los criterios (utilizando la técnica de diagrama de afinidad) teniendo como premisa que se seleccionen sólo aquellos que sean fáciles de comprender y de aplicar objetivamente por los alumnos que actuarán como evaluadores. Tanto la tormenta de ideas como el diagrama de afinidad son dos técnicas incluidas en el programa de la asignatura como sesiones prácticas. El profesor se encarga de sintetizar los criterios y de incluir una guía para facilitar la puntuación.

Tres semanas antes de la exposición, se presentan a los alumnos los criterios definitivos (ver Apéndice B), que cada alumno utiliza para evaluar a 3 compañeros. Además, se cuelgan en red las grabaciones en vídeo de las presentaciones realizadas durante la asignatura. Estas presentaciones son independientes de la actividad objeto de este trabajo y consisten en la actuación de diferentes alumnos (casi el 50% de los asistentes) como portavoces de las actividades de grupos que se han realizando durante las sesiones de clase. El objetivo principal es que los alumnos se evalúen a sí mismos y a otros dos compañeros utilizando los criterios de puntuación. De este modo reciben realimentación de su habilidad como oradores y se familiarizan con el uso del instrumento de evaluación. En sesiones posteriores se da la oportunidad de que los alumnos comenten las dificultades surgidas durante la utilización de los criterios.

Durante las dos semanas siguientes, los alumnos trabajan por su cuenta realizando las entrevistas. El día de la presentación, los alumnos cuentan con una hora de clase para preparar sus presentaciones de tres minutos como máximo cada una.

Durante la segunda hora de clase proceden a realizar las presentaciones, puntuadas por sus compañeros y por el profesor.

Con el fin de no aislar a los alumnos de los contenidos de las presentaciones, éstos evalúan sólo una de cada cuatro presentaciones. La asignación de la exposición que deben evaluar se hace en función del lugar que los alumnos ocupan físicamente en el aula, por lo que se puede considerar una asignación al azar. El profesor evalúa todas las presentaciones y utiliza la misma parrilla de criterios que los alumnos.

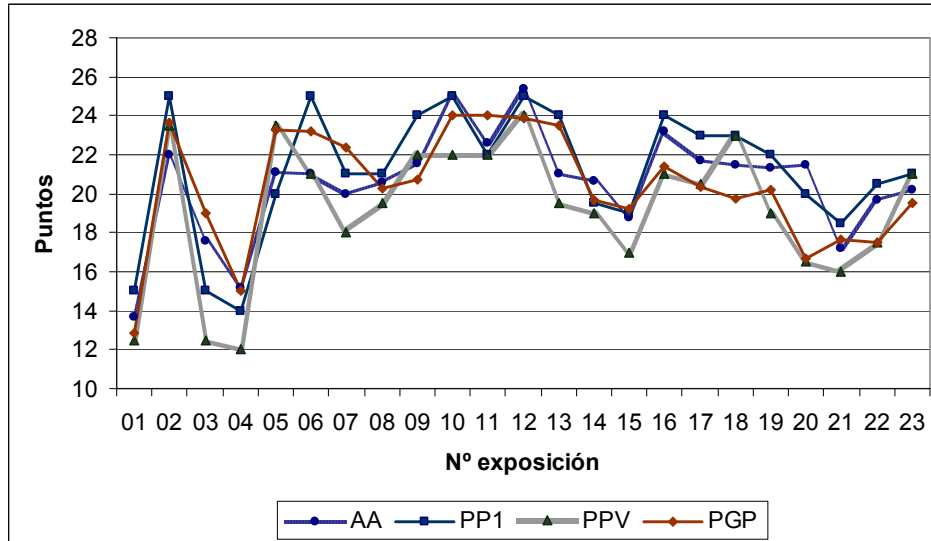
Cuatro meses después de la exposición, un grupo de cuatro profesores se reúne con el profesor de la asignatura para evaluar las presentaciones orales grabadas en vídeo. De los cuatro profesores, uno imparte la misma asignatura, pero en un centro distinto, otro procede de la misma área de conocimiento y los otros dos pertenecen a áreas de conocimiento distintas.

En total se recogen cuatro conjuntos de datos:

- Las puntuaciones que dan los alumnos a las presentaciones orales de sus compañeros (cada exposición se evalúa por 7-12 alumnos) (AA).
- Las puntuaciones del profesor de la asignatura de las presentaciones orales de los alumnos (PP1).
- Las puntuaciones del profesor de la asignatura de las presentaciones orales grabadas en vídeo (PPV).
- Las puntuaciones de cuatro profesores de las presentaciones orales grabadas en vídeo (PGP).

3. Análisis de resultados

La figura 1 muestra el gráfico de las puntuaciones obtenidas en las diferentes presentaciones orales llevadas a cabo. Se calcula el promedio en aquellas series de datos donde se dispone de más de una puntuación por exposición (AA y PGP). A simple vista, se puede observar que existe bastante similitud entre los diferentes datos; una serie de pruebas estadísticas aporta rigor a esta interpretación.



AA: promedio de las puntuaciones de los alumnos a las presentaciones orales.
 PP1: puntuación del profesor de la asignatura a las presentaciones orales.
 PPV: puntuación del profesor a las presentaciones orales grabadas en vídeo.
 PGP: promedio de las puntuaciones de 4 profesores a las presentaciones orales grabadas en vídeo.

Figura 1. Puntuaciones obtenidas en las presentaciones orales

En primer lugar, se han calculado los estadísticos descriptivos y se han comprobado que las cuatro series de datos (AA, PP1, PPV y PGP) presentan una distribución normal, mediante la Prueba de Kolmogorov-Smirnov para una muestra (VVAA, 2002). En Tabla 1 se presentan las correlaciones de Pearson entre las variables (AA, PP1, PPV y PGP).

	N	Media	Desviación típica	Mínimo	Máximo	AA	PP1	PPV	PGP
AA: promedio de las puntuaciones de los alumnos que evalúan esa presentación	242	20,538	2,7155	13,7	25,4	1	,875(**)	,856(**)	,811(**)
PP1: puntuaciones del profesor de la asignatura el día de la presentación	23	21,152	3,2768	14,0	25,0	-	1	,878(**)	,791(**)
PPV: puntuaciones del profesor de la asignatura sobre vídeo	23	19,239	3,5543	12,0	24,0	-	-	1	,808(**)
PGP: promedio de las puntuaciones de los otros 4 profesores sobre vídeo	87	20,326	3,0168	12,8	24,0	-	-	-	1

** La correlación es significativa al nivel 0,01 (bilateral).

Tabla 1. Estadísticos descriptivos

Las puntuaciones realizadas por el profesor de la asignatura de las presentaciones orales grabadas en vídeo (PPV) están altamente correlacionadas con el promedio de las puntuaciones de los otros cuatro profesores (PGP) (0,808). Las diferencias entre las medias de ambas puntuaciones son significativas con una prueba T ($\alpha = 0,02$). Esto indica que, en promedio, las puntuaciones del profesor de la asignatura son 1,08 puntos menores que las de los otros profesores. En total hay 4 presentaciones (17%) en las que la diferencia entre las puntuaciones de la variable PPV y las de la variable PGP son mayores que 3 puntos (una desviación estándar de la variable PGP).

De manera análoga, la correlación entre el promedio de las puntuaciones de los alumnos (AA) y la del profesor de la asignatura, el día de las presentaciones (PP1) es altamente significativa (0,875). Sin embargo, las diferencias entre las medias de ambas puntuaciones no son significativas con una prueba T. En total, sólo hay una exposición (4%) en la que la puntuación AA se diferencia de PP1 en más de una desviación estándar de las puntuaciones PP1 (3,27 puntos).

Por lo tanto, respecto a las dos primeras preguntas de nuestra investigación podemos concluir que las puntuaciones de un único profesor coinciden bastante con las del grupo de cuatro profesores. La diferencia promedio entre ambas puntuaciones, aunque es significativa estadísticamente, sólo es de 1 punto sobre un total de 27. Por otra parte, en el 83% de los casos la diferencia entre las puntuaciones PPV y PGP es menor que 3 puntos sobre una escala de 27 puntos. No se dispone de datos de investigaciones similares que hayan estudiado estas variables, por lo que no existen referentes comparativos.

Del mismo modo, y de manera más acusada, los promedios de las puntuaciones realizadas por los alumnos a sus compañeros concuerdan, casi totalmente, con la puntuación del profesor de la asignatura. Los resultados obtenidos coinciden con estudios anteriores que encuentran altas correlaciones entre las puntuaciones de los alumnos y las de los profesores⁷. Los resultados obtenidos en la presente investigación también coinciden con todos los estudios que han observado que la dispersión de las puntuaciones de los alumnos es menor y que discriminan menos⁸, aunque en este caso las diferencias de dispersión son bastante pequeñas. Por último, las puntuaciones que dan los alumnos a sus compañeros son ligeramente menores que las que otorga el profesor.

⁷ Al-Fallay, 2004; Falchikov, 1995; Falchikov y Goldfinch, 2000; Freeman, 1995; Langan et al., 2005; MacAlpine, 1999; MacPherson, 1999; Magin y Helmore, 2001.

⁸ Cheng y Warren, 1999; Falchikov y Goldfinch, 2000; Freeman, 1995; Kwan y Leung, 1996; Magin y Helmore, 2001.

Una vez analizado el grado de acuerdo, pasamos a comentar los resultados relacionados con la fiabilidad de las puntuaciones de los alumnos y del profesor. En ambos casos se sigue un procedimiento análogo.

En primer lugar, se calcula la fiabilidad entre evaluadores (r_{nn}) del grupo de alumnos que puntúan a sus compañeros o del grupo de profesores que evalúan las presentaciones orales grabadas en vídeo, sin incluir la puntuación del profesor de la asignatura. Posteriormente se añaden a los análisis las puntuaciones del profesor de la asignatura para estimar su fiabilidad (r^*_{nn}). Por último, si la fiabilidad estimada del profesor de la asignatura (r_{tt}) es superior a la de los alumnos o la del conjunto de profesores colaboradores en la investigación (r_{11}), se calculan cuantos alumnos o profesores de otras disciplinas (DeltaN) sería necesario que evaluaran cada exposición para que el promedio de sus puntuaciones tuviese una fiabilidad similar a la del profesor de la asignatura.

Los resultados de estos análisis (ver Tabla 2) no pueden ser más elocuentes. La fiabilidad de las puntuaciones de los alumnos ($r_{nn} = 0,90$) es muy elevada cuando interviene una cantidad grande de evaluadores (en este caso más de 10 por cada presentación oral). Sin embargo, la fiabilidad estimada de las puntuaciones de un único alumno evaluador es un poco baja ($r_{11} = 0,47$), pero prácticamente igual cuando se compara con un evaluador de los 4 profesores que evalúan las grabaciones.

	Puntuación de los alumnos a los compañeros	Puntuación de los profesores que evalúan las presentaciones grabadas en vídeo
Número de evaluadores	43	4
Numero de presentaciones evaluadas	23	23
Numero de observaciones	242	87
Promedio de evaluadores por exposición (N)	10,52	3,78
Estadístico ANOVA de las puntuaciones de n evaluadores (F)	10,34	4,20
Estadístico ANOVA cuando se añaden las puntuaciones del profesor de la asignatura a las puntuaciones de los n evaluadores (F*)	12,24	6,13
Fiabilidad de las puntuaciones de N evaluadores (r_{nn})	0,90	0,76
Estimación de la fiabilidad de un solo evaluador (r_{11})	0,47	0,46
Fiabilidad de las puntuaciones cuando se incorporan las del profesor de la asignatura a los n evaluadores (r^*_{nn})	0,92	0,84
Estimación de la fiabilidad de las puntuaciones del profesor de la asignatura (r_{tt})	0,66	0,66
Estimación del número de alumnos/profesores que deberían puntuar cada exposición para que su fiabilidad fuese similar a la del profesor de la asignatura (DeltaN)	2,14	2,27

Tabla 2. Análisis de la fiabilidad de las puntuaciones de alumnos y profesores

Ahora bien, la fiabilidad estimada de las puntuaciones del profesor de la asignatura son sensiblemente mayores ($r_{tt} = 0,66$) y prácticamente constantes al contrastarla con la de las variables AA y PGP. Tanto la fiabilidad de las puntuaciones de los alumnos como la del profesor de la asignatura (r_{nn} , r_{11} y r_{tt}) son más elevadas que las obtenidas en la investigación de Magin y Helmore (2001). Sin embargo, la fiabilidad de las puntuaciones de los otros cuatro profesores es prácticamente igual a la obtenida en dicha investigación.

Existen varias razones que pueden explicar la alta fiabilidad encontrada. En primer lugar, el hecho de que sólo se han incluido criterios objetivos y que cada uno de los aspectos tiene una guía detallada para la puntuación. Se ha dejado de lado otro tipo de criterios (relacionados con aspectos más subjetivos) que también son importantes en la evaluación de presentaciones en público (por ejemplo si gusta, si capta la atención del oyente, si interesa, entre otros). Sin embargo, esta investigación se centra en fomentar la fiabilidad de las puntuaciones. Por otra parte, la parrilla de criterios se plantea como un instrumento formativo, ya que sirve como guía a los alumnos e indica las pautas a seguir en la exposición oral (por ejemplo hablar alto y claro o contacto visual con la audiencia).

En segundo lugar, tanto los alumnos como el profesor de la asignatura prueban previamente la parrilla y se familiarizan con los criterios y la forma de puntuación, los otros cuatro profesores que colaboran en la investigación conocen los criterios en la misma sesión que puntúan las presentaciones grabadas en vídeo. Antes de empezar esta sesión de puntuación, reciben una explicación de 15 minutos por parte del profesor de la asignatura, tras la cual tienen unos minutos para leer los criterios y consultar dudas. Una vez resueltas las dudas, comienza la puntuación de las presentaciones orales grabadas en vídeo. Esta puede ser una de las causas por la que la fiabilidad estimada de uno de los profesores (r_{11}) es sensiblemente menor a la del profesor de la asignatura (r_{tt}). Se considera que estos cuatro profesores tenían una experiencia similar y, en algunos casos, mucho mayor que el profesor de la asignatura en la evaluación de presentaciones orales. Por lo tanto, esa mejora de fiabilidad puede explicarse por la familiaridad con los criterios utilizados. Por otra parte, puesto que r_{11} es similar en los profesores (personas habituadas a puntuar a sus alumnos, pero no familiarizados con los criterios que utilizados en este caso) y en los alumnos (personas sin experiencia en evaluar a sus compañeros, pero entrenados en el uso de la parrilla de criterios) se intuye que los criterios utilizados y el entrenamiento previo paliaban la inexperiencia.

4. Conclusiones

En primer lugar, se advierte de que las conclusiones que aquí se presentan sólo son extrapolables a situaciones similares a la expuesta: presentaciones orales, evaluadas con criterios elaborados por los propios alumnos, entrenados para su uso. Cabe añadir que las puntuaciones de los alumnos contribuyen a la evaluación sumativa de sus compañeros, pero también, los alumnos reciben una puntuación como evaluadores, lo que les lleva a tomarse en serio la actividad. También se intenta motivar a los alumnos haciéndoles notar que esta actividad puede contribuir a que practiquen la evaluación del rendimiento de otras personas (subordinados o compañeros), una habilidad que es muy probable que necesiten en su vida profesional futura. Este procedimiento se ha extraído de las recomendaciones de investigaciones previas; sin ellas es probable que los resultados fuesen diferentes.

En estas condiciones se puede considerar que las puntuaciones de las presentaciones, calculadas como el promedio de los puntos recibidos de varios alumnos actuando como evaluadores, no se diferencian de manera significativa de las puntuaciones que otorga el profesor de la asignatura. Por ello, podemos utilizar las puntuaciones derivadas de la evaluación entre pares en la evaluación final de los alumnos sin que ello distorsione los resultados.

Es más, las puntuaciones del profesor de la asignatura presentan algunas diferencias al compararlas con el promedio de puntuaciones que propone el grupo de 4 profesores que ha colaborado en la investigación. Aunque esas diferencias no son grandes nos recuerdan que, por muy experimentado que sea un profesor, no podemos dar por sentado que las puntuaciones que éste otorga son absolutamente acertadas, ya que inciden otros muchos factores. Esta conclusión se corrobora con el dato de que las fiabilidades estimadas de un solo evaluador son moderadamente bajas en el caso del profesor de la asignatura y algo bajas en el caso de los alumnos o de los otros profesores participantes. Por lo tanto, lo más aconsejable, desde el punto de vista de la fiabilidad, sería utilizar las puntuaciones de varios evaluadores simultáneamente.

Los resultados de este estudio parecen indicar con bastante claridad que los alumnos pueden llegar a ser unos buenos evaluadores de las presentaciones orales de sus compañeros y que sus puntuaciones podrían servir como calificación de sus compañeros sin que se aprecien diferencias significativas con las puntuaciones que propone el profesor de la asignatura.

Si la evaluación de los alumnos es fiable, el profesor puede delegar parte de la tarea de evaluar en el alumno. De esta manera, al mismo tiempo que se fomenta la autonomía, responsabilidad y participación de los alumnos en el proceso de evaluación (aspectos que suelen originar una mayor motivación de los estudiantes), se consigue liberar tiempo en la ya de por sí sobrecargada agenda del profesor.

Si no se considera adecuado dejar toda la responsabilidad en manos de los alumnos, existen dos posibilidades: hacer intervenir a varios profesores o utilizar evaluación participativa. En el contexto de la asignatura *Gestión de Empresas* lo más habitual es no poder contar con la participación de más profesores. Por lo tanto, se podría mejorar la fiabilidad si cada exposición se puntuara por el profesor de la asignatura junto con varios alumnos (entre dos y cuatro). Con ello se podría obtener unos valores de r^*_m superiores a 0,80 que puede considerarse bastante adecuados (Magin y Helmore, 2001). Con este procedimiento, no existe un ahorro de tiempo, pero se mejoraría la fiabilidad de la evaluación y se conseguiría implicar a los alumnos en el proceso, con los beneficios formativos y de motivación que ello conlleva.

Cabe recalcar como factores importantes para fomentar un elevado grado de acuerdo en las puntuaciones una selección de criterios lo más objetivos posibles y la creación de guías para la puntuación, la familiaridad o entrenamiento con los criterios de evaluación y la experiencia en la evaluación de presentaciones orales. Cuanto más presentes estén estos factores, mayor será el grado de acuerdo de las puntuaciones. Es más, manejando adecuadamente los dos factores, selección de criterios y entrenamiento, que caen bajo el campo de actuación de un profesor, se pueden conseguir grados de acuerdo muy elevados, incluso con pocos evaluadores de cada presentación. Por lo tanto, no parece haber limitación, al menos por el aspecto de grado de acuerdo, para fomentar una mayor participación de los alumnos en la puntuación de calificación final de sus compañeros.

La forma en que se ha construido la parrilla de criterios para la puntuación, con la participación activa de los alumnos y las actividades para que los alumnos se entrenaran en el uso de la parrilla, también puede haber influido favorablemente en el grado de acuerdo de las puntuaciones. No obstante, no se está en condición de valorar su efecto en estos momentos. Se pretende continuar la investigación con alumnos de cursos posteriores que no han participado en la elaboración de la parrilla de criterios y ver los resultados que se producen. Otro aspecto a analizar sería qué pasa cuando los alumnos usan la parrilla en las mismas condiciones que el grupo de 4 profesores. Es decir, viendo las presentaciones grabadas en vídeo tras una explicación de los criterios durante 15 minutos sin entrenamiento en el uso de la parrilla de criterios.

Por último, hay que tener en cuenta que la implantación de la evaluación entre pares, y su utilización en la calificación final de los alumnos, se verá dificultada en entornos universitarios tradicionales, sobre todo si la transferencia de responsabilidades cambia el equilibrio de poder entre los profesores y los alumnos (Dochy et al., 1999; MacPherson, 1999; Tariq et al., 1998). Además, cuando se pone en marcha la autoevaluación o evaluación de pares, la gestión del proceso se complica y se invierte más tiempo. En palabras de Cheng y Warren (1999) se recomienda revisar, simplificar y mecanizar el procedimiento para no sobrecargar a los docentes.

Bibliografía

- Al-Fallay, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System* 32(3): 407-427.
- Cheng, W. y Warren, M. (1999). Peer and teacher assessment of the oral and written tasks of a group project. *Assessment & Evaluation in Higher Education* 24(3): 301-314.
- Dochy, F., Segers, M. y Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3): 331-350.
- Falchikov, N. (1995). Peer feedback marking - developing peer assessment. *Innovations in Education and Training International*, 32(2): 175-187.
- Falchikov, N. y Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3): 287-322.
- Freeman, M. (1995). Peer assessment by groups of group work. *Assessment & Evaluation in Higher Education*, (20): 289-300.
- Gatfield, T. (1999). Examining student satisfaction with group projects and peer assessment. *Assessment & Evaluation in Higher Education*, 24(4): 365-377.
- Kane, J. y Lawler III, E. E. (1978). Methods of peer assessment. *Psychological Bulletin*, 85(3): 555-586.
- Kwan, K.-P. y Leung, R. (1996). Tutor versus peer group assessment of student performance in a simulation training exercise. *Assessment & Evaluation in Higher Education*, 21(3): 205-214.
- Langan, M., Wheeler, P., Shaw, E., Haines, B., Cullen, R., Boyle, J., Penney, D., Oldekop, J., Ashcroft, C., Lockett, L. y Preziosi, R. (2005). Peer assessment of oral presentations: Effects of student gender, university affiliation and participation in the development of assessment criteria. *Assessment & Evaluation in Higher Education*, 30(1): 21-34.
- MacAlpine, J. M. K. (1999). Improving and encouraging peer assessment of student presentations. *Assessment & Evaluation in Higher Education*, 24(1): 15-25.
- MacLellan, E. (2001). Assessment for learning: The differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education*, 26(4): 307-318.

- MacPherson, K. (1999). The development of critical thinking skills in undergraduate supervisory management units: Efficacy of student peer assessment. *Assessment & Evaluation in Higher Education*, 24(3): 273-284.
- Magin, D. J. (2001a). A novel technique for comparing the reliability of multiple peer assessments with that of single teacher assessments of group process work. *Assessment & Evaluation in Higher Education*, 26(2): 139-152.
- Magin, D. J. (2001b). Reciprocity as a source of bias in multiple peer assessment of group work. *Studies in Higher Education*, 26(1): 53-63.
- Magin, D. J. y Helmore, P. (2001). Peer and teacher assessments of oral presentation skills: How reliable are they? *Studies in Higher Education*, 26(3): 287-298.
- Orsmond, P., Merry, S. y Reiling, K. (1996). The Importance of marking criteria in the use of peer assessment. *Assessment & Evaluation in Higher Education*, 21(3): 239-250.
- Orsmond, P., Merry, S. y Reiling, K. (1997). A study in self-assessment: tutor and students' perceptions of performance criteria. *Assessment & Evaluation in Higher Education*, 22(4): 357-370.
- Orsmond, P., Merry, S. y Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education*, 25(1): 23-38.
- Reynolds, M. y Trehan, K. (2000). Assessment: A critical perspective. *Studies in Higher Education*, 25(3): 267-278.
- Searby, M. y Ewers, T. (1997). An evaluation of the use of peer assessment in higher education: A case study in the School of Music, Kingston University. *Assessment & Evaluation in Higher Education*, 22(4): 371.
- Sullivan, K. y Hall, C. (1997). Introducing students to self-assessment. *Assessment & Evaluation in Higher Education*, 22(3): 289-305.
- Tariq, V. N., Stefani, L. A. J., Butcher, A. C. y Heylings, D. J. A. (1998). Developing a new approach to the assessment of project work. *Assessment & Evaluation in Higher Education*, 23(3): 221-240.
- VVAA. (2002). *SPSS 10 Guía para el análisis de datos*. Servicio de Informática Universidad de Cádiz.
- Ward, M., Gruppen, L. y Regehr, G. (2002). Measuring self-assessment: Current state of the art. *Advances in Health Sciences Education*, 7(1): 63-80.

Apéndice A. Cálculo de la fiabilidad.

Adaptado de Magin (2001a) y Magin y Helmore (2001)

Para el cálculo de la fiabilidad de las puntuaciones de diversos evaluadores, podemos usar los resultados del Análisis de la Varianza de un factor. Los datos que necesitamos son el número de evaluadores promedio por exposición (N) y el ratio F (suma de cuadrados entre grupos de evaluadores, dividido por la suma de cuadrados dentro del grupo de evaluadores). Ambos datos son proporcionados en las tablas de resumen ANOVA de cualquier programa informático que realice estos cálculos estadísticos. Denotaremos por F* cuando se añadan, para el cálculo de los ANOVA, las puntuaciones del profesor de la asignatura a las puntuaciones de los N evaluadores.

Las variables de la investigación han sido calculadas del siguiente modo:

- Fiabilidad de las puntuaciones dentro del grupo de N evaluadores:
 - o $r_{nn} = (F-1)/F$
- Estimación de la fiabilidad de un solo evaluador:
 - o $r_{11} = (F-1)/(F+N-1)$
- Fiabilidad de las puntuaciones dentro del grupo cuando se incorporan las puntuaciones del profesor de la asignatura a los N evaluadores:
 - o $r_{nn}^* = (F^*-1)/F^*$
- Estimación de la fiabilidad de las puntuaciones del profesor de la asignatura:
 - o $r_{tt} = (F^*-F)/(F^*-F+1)$
- Estimación del número de evaluadores que deberían puntuar cada exposición para que su fiabilidad fuese similar a la del profesor de la asignatura:
 - o $\text{DeltaN} = N(F^*-F)/(F-1)$

Apéndice B. Plantilla para la puntuación de las presentaciones orales

Aspectos	Puntuación											
	0	1	2	3	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Exp 6	Exp 7	
Mirada	Sólo lee notas o transparencia	Mira a una sola persona o a unos pocos	mira a todos ojos a todos alguna vez	mira a todos a los ojos con frecuencia								
Naturalidad	No	Poco	---	Si								
Al hablar	No se oye	Se oye pero habla demasiado rápido/lento	Se oye pero habla un poco rápido/lento	Se oye y ritmo adecuado								
Transparencias legibles	Es bastante difícil leerlas	Cuesta un poco	--	Si								
Participan las dos personas en la exposición	No	Se nota mucha diferencia entre una y otra persona	--	Si								
Subtotal A												
Transparencias	No incluye dibujos ni colores	Tiene más de un color	Incluye gráficos o dibujos	Colores, gráficos y dibujos								
Transparencia esquematizada	No esquematizada Párrafos largos	Esquematizada, más de 14 líneas de texto	--	14 líneas de texto o menos por transparencia								
Terminan antes de los 3 minutos	El profesor les tiene que avisar	--	--	Si								
Exposición centrada en aspectos de la asignatura	El tema no pertenece a la asignatura o no cumple los requisitos de la actividad	--	--	Si								
Subtotal B												
Puntos totales: Subtotal A+ Subtotal B (máximo 27 puntos)												

Investigación cualitativa y cuantitativa en la coevaluación: correlación profesor-alumnos

Begoña Montero Fleta

Esta investigación de carácter cualitativo y cuantitativo se centra en la evaluación compartida, en línea con los trabajos de evaluación del trabajo en grupo de Freeman (1995), el análisis de los criterios de corrección en la evaluación por pares investigada por Orsmond, Merry y Reiling (1996) y las diferencias de la puntuación otorgada por el profesor y los alumnos observadas por Stefani (1994), Kwan y Leung (1996) y MacAlpine (1999). El estudio que aquí se presenta aborda, por una parte, tanto el proceso como el producto final a partir de la presentación oral de una experiencia de trabajo en grupo llevada a cabo en un entorno universitario y su posterior coevaluación y autoevaluación del alumno con respecto al proceso. Por otra parte, se analiza la coevaluación de otros alumnos junto con el profesor del producto final. Los resultados estudian la correlación profesor-alumno y se contrastan con los obtenidos en otras investigaciones.

Palabras clave: *presentaciones orales, coevaluación, acuerdo entre jueces, evaluación entre pares, evaluación del proceso y del producto, validez y fiabilidad.*

1. Introducción

El presente trabajo supone una innovación en la utilización de metodologías y programas específicos orientados a desarrollar un modelo de aprendizaje que haga más partícipes a los estudiantes. Su puesta en práctica implica un cambio de paradigma que pasa de una educación centrada en la docencia de los profesores, a otra donde el centro de atención es el alumno y su aprendizaje, siendo los docentes verdaderos facilitadores de un nuevo tipo de aprendizaje con aplicación directa al futuro profesional del discente. Este aprendizaje para toda la vida (*life long learning*) es

un elemento esencial para alcanzar una mayor competitividad, para mejorar la cohesión social, la igualdad de oportunidades y la calidad de vida (Delors et al., en Edwards y Llopis, 2004).

La investigación abordada en este trabajo se centra en la puesta en práctica de un trabajo en grupo que supone una reforma metodológica a nivel de enseñanza-aprendizaje por la implicación directa del alumno en la valoración del sistema y práctica de la evaluación. La presente investigación tiene como antecedentes las experiencias de evaluación del trabajo en grupo de Freeman (1995), el análisis de los criterios de corrección en la evaluación compartida investigada por Orsmond, Merry y Reiling (1996) y las diferencias de la puntuación otorgada por el profesor y asignada por el alumno de Kwan y Leung (1996) y MacAlpine (1999). Este análisis está basado en una experiencia llevada a cabo en la Facultad de Informática de la Universidad Politécnica de Valencia (UPV) con alumnos de Documentación, especialidad de segundo ciclo, en la asignatura de carácter obligatorio *Inglés Oral Avanzado*, cursada por alumnos con un dominio de lengua heterogéneo. El trabajo propuesto se desarrolla en grupos de tres alumnos elegidos por ellos mismos, respondiendo a lazos de amistad, afinidad de horarios, etc. La actividad central comprende la preparación de una exposición oral de 10 minutos de duración que describe la experiencia personal de cada grupo acerca de su participación en un foro de discusión en Internet, donde se tratan temas directamente relacionados con aspectos concretos de su campo de estudio. El trabajo se asigna y se perfila el primer día de clase, de forma que los alumnos comiencen su intervención en el foro de forma inmediata.

2. Objetivos

Conscientes de que la evaluación debe ser un proceso participativo y cooperativo donde se implican todos aquellos aspectos personales que intervienen, los objetivos marcados son triples, pues persiguen:

1. La implicación responsable del alumno en el trabajo en grupo (autoevaluación del alumno en el proceso).
2. La participación del alumno en varios ámbitos de la evaluación:
 - la evaluación del trabajo individual de cada uno de los componentes de su propio grupo (evaluación de sus pares con respecto al proceso)
 - la evaluación del trabajo presentado por compañeros fuera del grupo (evaluación de sus pares con respecto al producto).

Esta evaluación del proceso y producto propicia una reorganización conceptual del sistema educativo para adaptarse a los modelos de formación continuada a lo largo de toda la vida en la línea señalada por Milton (2005), quien defiende que los estudiantes pueden desarrollar destrezas evaluativas que les permite analizar su propio trabajo y el de los demás.

3. La comprobación del acuerdo entre jueces mediante la observación de las posibles diferencias entre la puntuación asignada alumno-alumno y profesor-alumno en una actividad concreta. La hipótesis nula (H_0) que se formula es que no existen diferencias entre las calificaciones otorgadas por el alumno y las del profesor. La hipótesis alternativa pretende demostrar que estas diferencias sí que existen y que no se deben al azar. El intervalo de confianza que se establece es del 95%.

3. Materiales y método

La actividad asignada para alcanzar los objetivos propuestos se basa en la participación de cada grupo de alumnos en un foro de discusión en Internet en el campo de la Biblioteconomía y Documentación. Los detalles y conclusiones de dicha experiencia se presentan de forma oral en clase, teniendo en cuenta que el hecho de que el trabajo propio sea contemplado por unos “espectadores”, en este caso los propios compañeros, es la mejor manera de motivar al alumno, no sólo en el cuidado del contenido de la presentación sino también en la pulcritud de su exposición (Durán, Sánchez-Reyes y Beltrán, 2004).

La evaluación del proceso corre a cargo de los alumnos con respecto a sus compañeros integrantes de grupo, actividad ésta que aporta información válida con respecto a la tarea realizada e implicación personal de los participantes, que en ocasiones escapa al profesor:

You are unlikely to be able to accurately judge the individual contributions made or to adequately assess the group process itself. In this situation it is the students themselves who are in the best position to appraise the relative contributions of individual team members to the group work they have undertaken together (Sherpa: 2004:10).

La evaluación del producto se lleva a cabo tanto por el profesor como por los alumnos. La presentación de las tres personas de cada grupo se evalúa por otros tres compañeros elegidos por el profesor al azar. Tabla 1 muestra el proceso global de evaluación del grupo efectuada:

PROCESO	PRODUCTO
- autoevaluación de los alumnos - coevaluación de los alumnos de un mismo grupo	- evaluación del profesor - coevaluación de 3 alumnos no pertenecientes al grupo (anónima)

Tabla 1. Evaluación trabajo en grupo

El trabajo se desarrolla en varias fases:

- a) La fase de preparación es decisiva para el correcto funcionamiento de toda actividad docente. En este caso comprende la organización del trabajo a realizar y el nivel de implicación del grupo. Ello supone la localización e inscripción en un fórum de discusión en Internet cuyo temática se relacione directamente con contenidos de sus estudios y futura profesión. A continuación se delimita el método de trabajo a seguir, acordando frecuencia, tiempo y lugar de reuniones.

Con el fin de lograr una implicación total de los alumnos en el proceso de evaluación se comienza por consensuar con ellos en una primera fase de la experiencia los criterios referentes a su grado de participación en el grupo (evaluación del proceso) y se diseñan las parrillas de valoración que tendrían en cuenta aspectos tales como la aportación de ideas, la implicación en el foro y la participación en la preparación de la presentación (ver Apéndice A).

Para evaluar la presentación oral (evaluación del producto) se define la parrilla con criterios detallados que determinan la calidad del trabajo oral valorando tanto el rigor en la exposición, contenido, organización de la información y apoyo visual como la corrección lingüística (ver Apéndice B). Para conseguir una mayor objetividad en las evaluaciones se reflexiona y se hace reflexionar sobre cuestiones de ética, precisión y fiabilidad, transparencia, rigor o proceso crítico en la tarea evaluadora, potenciando aspectos positivos y negativos así como fidelidad a los objetivos del trabajo.

Con el fin de que los alumnos dispongan de elementos de juicio claros acerca de la evaluación de una presentación se opta por deducir las exigencias de una buena presentación oral tomando como punto de referencia presentaciones orales comerciales (por ejemplo, Goodale, 1998) y grabaciones de otras presentaciones efectuadas en clase por alumnos de convocatorias

anteriores. El material seleccionado muestra errores típicos que se producen al exponer la información ante una audiencia. De esta forma:

- Se reflexiona sobre puntos débiles y puntos fuertes de las presentaciones en lo que se refiere tanto a su contenido como a su puesta en escena.
- Se determina la distribución de la información en las distintas partes de la presentación.
- Se estudian los aspectos lingüísticos característicos de cada una de sus partes.
- Se resumen normas a seguir en el diseño de las imágenes de PowerPoint referentes al tamaño de la fuente, o cantidad de información a incluir, ya que los alumnos tienden a sobrecargar las imágenes con información.

- b) En una fase prácticamente simultánea en el tiempo a la anterior se comienza la intervención en el foro, lo cual implica decisión acerca del contenido de los mensajes, redacción de los mismos, el intercambio de mensajes y análisis del flujo de mensajes generado. Esta actividad se desarrolla en el laboratorio en sesiones de prácticas de la asignatura. Desde el principio los alumnos cuentan con la colaboración del profesor para resolver dudas y problemas que pueden surgir, tales como dificultades de comunicación en algún foro o redacción de mensajes.
- c) La siguiente fase es la redacción del documento que resume la experiencia de participación en el foro de discusión y que sirve de base a la presentación oral. El documento se revisa por el profesor en presencia de los autores.
- d) Seguidamente se prepara la presentación oral. Como actividad individual de cada grupo con el profesor, cada alumno lee personalmente la información que va a exponer con el fin de corregir errores de pronunciación y entonación. Se hace especial hincapié en recordar que los parámetros de la evaluación de la presentación no se ciñen exclusivamente a aspectos de carácter lingüístico (ver Apéndice B). Se revisa igualmente la información recogida en las imágenes de PowerPoint.
- e) A continuación se hace la presentación oral en clase de la experiencia, actividad que para muchos supone un reto importante.
- f) Finalmente se procede a la evaluación de la experiencia de la forma siguiente:
1. Coevaluación o evaluación de los alumnos entre los miembros del grupo y la propia autoevaluación del alumno con respecto al proceso.

2. Coevaluación de otros alumnos externos al grupo del producto final.
3. Evaluación del profesor centrada también en el producto.

Para motivar la participación del alumno en el trabajo en grupo y en la evaluación se le asigna al proceso 1 punto de la calificación final, a repartir según criterio de los tres integrantes de cada grupo y en función de su grado de participación individual en la actividad. Para la evaluación de la presentación oral (evaluación producto) se asigna 1 punto de la calificación final individual de cada alumno, que se otorga en un 70% por el profesor y en un 30% por la media de la coevaluación por otros tres compañeros externos al grupo.

La investigación se completa con el estudio comparativo de los resultados obtenidos con respecto al producto con los obtenidos por Kwan y Leung (1996) MacAlpine (1999) en sus estudios encaminados a evaluar las presentaciones orales de los alumnos por alumnos y profesor.

4. Resultados

Los resultados muestran una aceptación de la actividad por los alumnos, tanto en su participación en el trabajo en grupo como en la presentación oral. La falta de práctica en hablar en público y la inseguridad que produce la comunicación en una lengua que no es la materna no son impedimento para que la mayoría de los alumnos superen la actividad con éxito, probablemente, como resultado del tiempo e interés dedicado a la preparación de la misma.

Con respecto a la evaluación del proceso cabe destacar que el 90% del alumnado opta por repartirse equitativamente el punto que se otorga, lo que implica una igual participación en cada uno de los aspectos que se juzgan. El 10% restante presenta muy ligeras variaciones. Con respecto a la evaluación del producto se tienen en cuenta los siguientes parámetros: exposición, contenido, organización de la información y dominio de la lengua.

En este estudio descriptivo se observan las medias y las desviaciones típicas de las evaluaciones efectuadas así como las calificaciones mínimas y máximas otorgadas con el fin de comprobar el acuerdo entre jueces siguiendo las directrices del estudio llevado a cabo por Kwan y Leung (1996). Para estos autores la media del acuerdo entre los distintos examinadores se observa analizando la media y la desviación típica de ambos; la media de la nota de los alumnos cae en el intervalo comprendido entre la media +/- una desviación estándar de las calificaciones otorgadas por el profesor.

	N	Mínimo	Máximo	Media	Desv. típ.
Profesor	37	4,00	10,00	7,8216	1,5964
Alumnos	37	6,00	9,00	7,4270	,8211
N válido (según lista)	37				

Tabla 2. Datos estadísticos descriptivos

Como muestra la tabla 2 se observan diferencias entre las puntuaciones máximas y las mínimas, es decir, entre el rango o recorrido. En el caso del profesor el rango es de 6 (10-4), mientras que en los alumnos es de 3 (9-6), lo que parece indicar que el alumno al tener menos experiencia en tareas de evaluación tiende a dar unas calificaciones más similares, lo que indica una menor discriminación.

Para determinar el grado de acuerdo entre profesores y alumnos se utilizan diferentes medidas: T-test como prueba de contraste de hipótesis; hipótesis nula: no existe diferencia y si existe es debida al azar; hipótesis alternativa: existe diferencia. Para comprobar que la variable tiene una distribución normal se realiza previamente el test de Kolmogoroff- Smirnov.

T-test para muestras relacionadas:

	N	Correlación	Sig.
Par 1 Alumno y profesor	37	,494	,002

Tabla 3. Correlaciones de muestras relacionadas

	Diferencias relacionadas						t	gl	Sig (bilateral)
	Media	Desviación típ	Error típ. de la media	Intervalo de confianza para la diferencia					
				Inferior	Superior				
Par 1 Aumno y Profesor	-,47568	1,40028	,23020	-,94255	-,00880	-2,066	36	,046	

Tabla 4. Prueba de muestras relacionadas

Como punto de corte para decidir aceptar o rechazar la hipótesis alternativa se utiliza el valor de $P = 0,05$; es decir, si la probabilidad de que un determinado resultado se debe al azar es menor del 5% se acepta que no es debido al azar,

admitiendo que en un 5% de los casos existe probabilidad de error. En el T-test se observa un nivel de significación estadística $P = 0,046$, luego $<0,05$, por tanto, rechaza la hipótesis nula y acepta la hipótesis alternativa (existe diferencia).

En la tabla 5 se muestra la correlación existente en la puntuación de cada una de las variables. Como se puede apreciar existe una muy alta correlación entre las distintas variables lo que confirma una alta correlación entre sí.

	Profesor					Alumno			
	Exp	Cont	Org	Leng		Exp	Cont	Org	Leng
Exp		0,80	0,71	0,86	Exp		0,74	0,86	0,95
Cont	0,80		0,91	0,83	Cont	0,74		0,77	0,74
Org	0,71	0,91		0,75	Org	0,86	0,77		0,89
Leng	0,86	0,83	0,75		Leng	0,95	0,74	0,89	

Exp.: Exposición Org.: Organización
 Cont.: Contenido Leng.: Lenguaje

Tabla 5. Coeficiente de correlación entre los distintos parámetros evaluados

La figura 1 recoge las correlaciones existentes entre las notas otorgadas por el profesor y la media de las puntuadas por los alumnos. Cada punto hace referencia a un alumno. El eje de ordenadas representa la media de las puntuaciones asignadas de forma anónima por 3 alumnos. El eje de abscisas indica la puntuación otorgada por el profesor. La recta de regresión lineal representa la correlación obtenida:

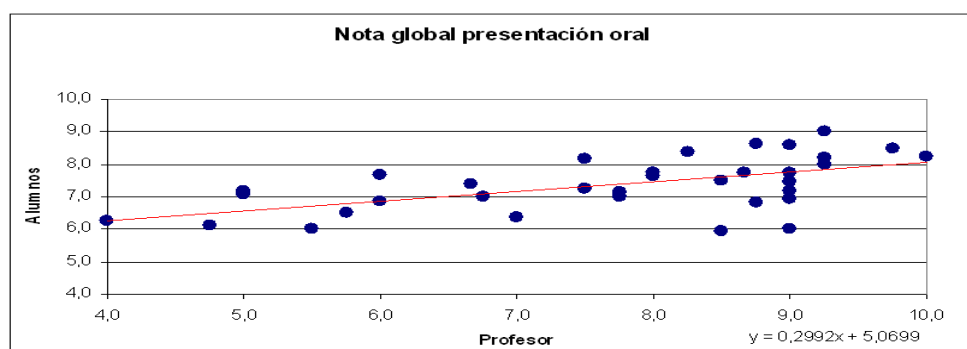


Figura 1. Relación entre puntuaciones otorgadas por el profesor y medias de las puntuaciones asignadas por alumnos

El coeficiente de correlación total ha resultado 0,58 y el gradiente de la regresión lineal calculado es 0,29, en línea con los resultados obtenidos por MacAlpine (1999). Como se observa, los alumnos son más exigentes que el profesor en la

puntuación otorgada; aunque hay que señalar que en los casos en que el profesor asigna una puntuación alta, la puntuación media concedida por los alumnos es mucho más baja, pero en los casos en que el profesor otorga una puntuación baja, ningún alumno coincide con ella y otorga una puntuación considerablemente más alta con el deseo de "salvar" al compañero. Estos resultados coinciden con los obtenidos por Stefani (1994), Freeman (1995), Kwan y Leung (1996) y MacAlpine (1999). La aparente falta de coherencia en estas ocasiones tiene una doble lectura. Por una parte, la falta de experiencia en contribuir en tareas de evaluación, que se puede mostrar en la dificultad en apreciar todos los aspectos a evaluar. Por otra parte, es evidente el falso espíritu de compañerismo que lleva a no otorgar en ocasiones una nota inferior al aprobado aun en casos de bloqueo del alumno en la exposición oral. Estos datos coinciden con los obtenidos por MacAlpine (1999) en un experimento de similares características.

Pero no todos los puntos evaluados reflejan la misma correlación existente entre la puntuación otorgada por el profesor y la otorgada por el alumno. En Figura 2 se muestran los resultados obtenidos en cada una de las variables evaluadas.

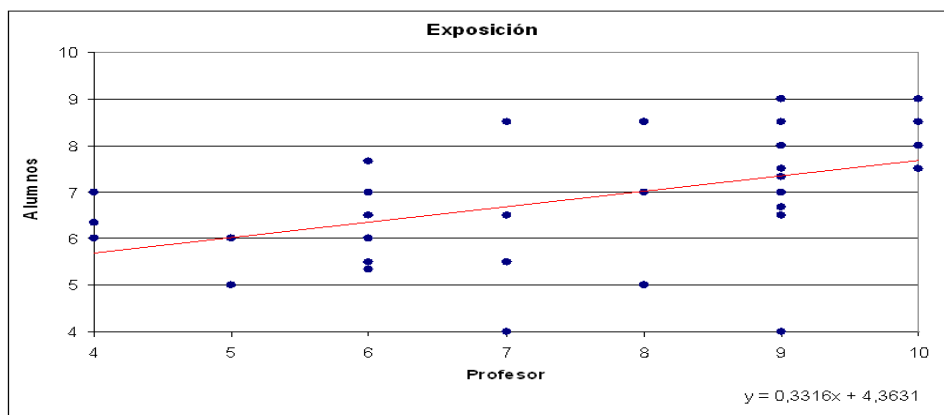


Figura 2. Exposición oral: relación entre notas otorgadas por el profesor y medias de las notas asignadas por alumnos

- El coeficiente de correlación en la exposición oral (volumen, contacto visual con la audiencia, naturalidad, velocidad de habla, posturas y movimientos) es 0,47. Como se observa en Figura 3, existe cierta dispersión en la muestra. Véase, por ejemplo, las diferencias en las exposiciones puntuadas con 9 por el profesor y sus correspondientes puntuaciones otorgadas por los alumnos (4, 6,5, 6,6, 7, 7,4, 7,5 8, 9). Los alumnos son más críticos en puntuar este aspecto como veremos a continuación. Es este parámetro en el que, aun sin

tener plenamente interiorizadas y asimiladas las claves específicas de una buena exposición, todos se ven capacitados para juzgar con criterios estrictos a su compañero. Los alumnos instintivamente establecen un paralelismo de la actuación a evaluar con otras habituales de transmisión de información ya sea en entornos académicos, como pueden ser la de sus profesores, o en contextos extraacadémicos como las de profesionales de medios de comunicación.

- En cuanto al contenido (claridad en la idea principal, tema adecuado para el tiempo disponible, tema adecuado para la audiencia, tema desarrollado con detalles, hechos y ejemplos importantes, tratamiento del tema en el tiempo asignado) se observa la correlación 0,64, la más alta de las distintas áreas puntuadas, tal y como nos muestra la figura 3. Es este el parámetro que muestra mayor acuerdo entre jueces, hecho que no sorprende pues comprende aspectos muy objetivos como, por ejemplo, si se ha completado la presentación en el tiempo asignado.

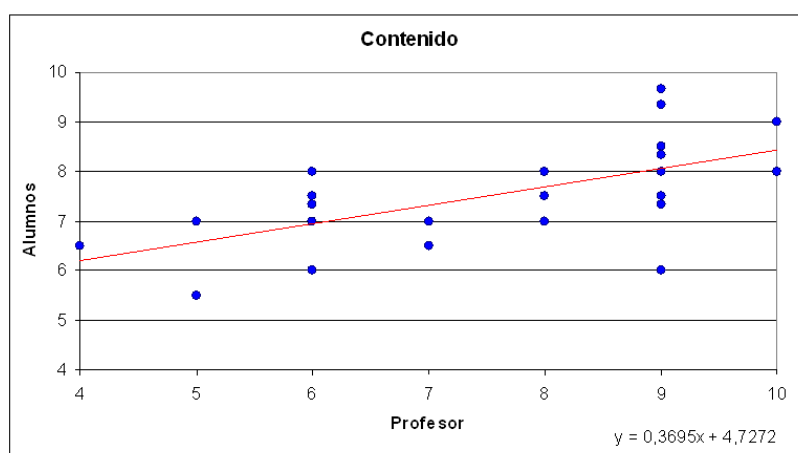


Figura 3. Contenido: relación entre notas otorgadas por el profesor y medias de las notas asignadas por alumnos

- En la puntuación de la organización de la presentación (introducción, transiciones, puntos principales claramente resaltados, desarrollo lógico de ideas, conclusión) se refleja que la correlación es 0,44 como muestra la figura 4.

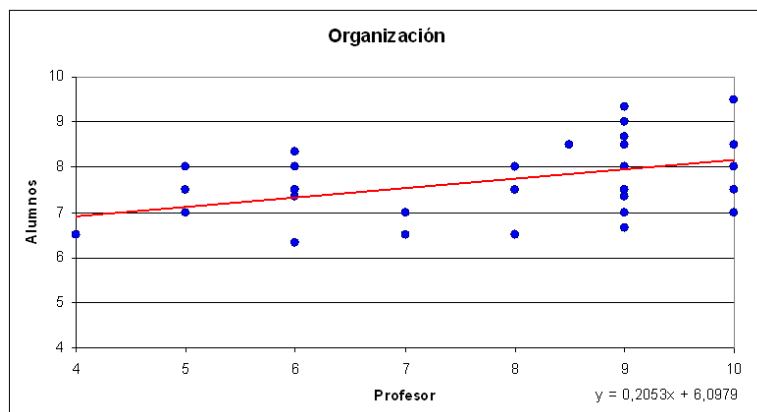


Figura 4. Organización: relación entre notas otorgadas por el profesor y medias de las notas asignadas por alumnos

- En este parámetro la correlación es menor puesto que no hay tanta dispersión en las puntuaciones otorgadas. Este hecho no debe sorprender pues su evaluación exige mayor experiencia; en palabras de los propios alumnos *'si atiendes bien a la puntuación a otorgar en unos aspectos te falta tiempo para observar otros'*, y los alumnos requieren más práctica para adquirir mayor fluidez en lograr que los puntos a juzgar en este parámetro sean más objetivables.
- En cuanto a la lengua (precisión en la comunicación, vocabulario apropiado para la audiencia, pronunciación/entonación, variedad de expresiones) la correlación existente es 0,47, como se observa en Figura 5.

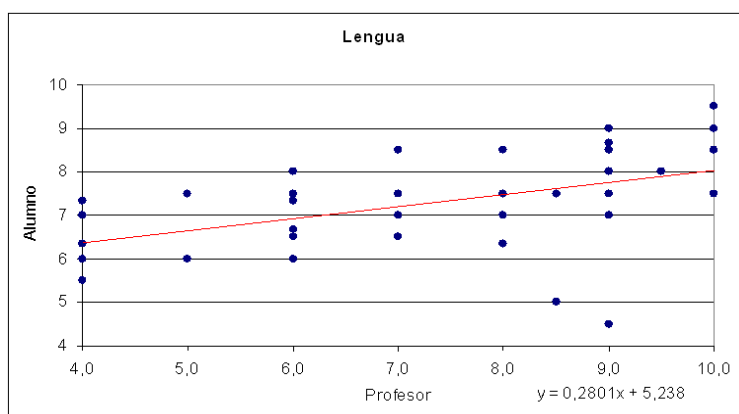


Figura 5. Lengua: relación entre notas otorgadas por el profesor y medias de las notas asignadas por alumnos

Se observa de nuevo que la puntuación media del profesor es superior con respecto a la media de los alumnos, quienes aparentemente acaban fijándose en el momento de puntuar en la aparente “fluidez, soltura, acento y facilidad en el manejo de la lengua” que, según los alumnos es lo que les da idea del dominio de la lengua.

5. Conclusión

El trabajo en grupo en toda su dimensión pedagógica nos ofrece una alternativa interesante que fomenta la colaboración y coordinación de los alumnos para llevar a cabo la tarea asignada. Igualmente, la evaluación compartida contribuye de forma positiva al rendimiento de los alumnos y les proporciona experiencia de aprendizaje en juzgar el trabajo realizado por compañeros de forma objetiva, tarea que será útil en su futuro profesional. La actividad que se presenta, consensuada desde el principio con los alumnos, resulta constructiva, participativa y entrena aspectos de mejora futura que incluya una propia autoevaluación del alumno con respecto a su propia exposición oral. La experiencia resulta gratificante para los alumnos en varios sentidos: participar activamente en un foro profesional resulta motivador, al poder hacer uso de los conocimientos adquiridos y poderlos aplicar en un entorno fuera del propiamente académico como puede ser Internet. Para muchos la presentación oral supone un reto importante que demuestra su capacidad de enfrentarse a exposiciones orales sobre temas profesionales en una lengua que no es la propia tras una concienzuda preparación del tema. La evaluación del proceso supone sólo un primer paso, pese a los sentimientos de compañerismo, distinguir que es justo que se reconozca el grado de aportación individual de cada integrante del grupo. La evaluación del producto constituye una práctica real encaminada a desarrollar la habilidad del alumno en valorar su propio trabajo y los aspectos débiles y fuertes de manera realista. Los resultados obtenidos han corroborado las aportaciones de Cheng y Warren (1999) y MacAlpine (1999).

La validez y fiabilidad de la evaluación del alumno son los aspectos a determinar en el presente estudio: si los alumnos son competentes al evaluar a sus compañeros junto al profesor. El contraste de hipótesis nos lleva a concluir que se pueden apreciar diferencias significativas entre las calificaciones otorgadas por profesor y alumnos. Esta investigación muestra que la tendencia de los alumnos es otorgar, en general, notas más bajas que el profesor en el caso de presentaciones con puntuaciones altas a juicio del profesor; sin embargo, el alumno que no alcanza los objetivos mínimos recibe por parte de los demás alumnos una nota superior a la del profesor. Los alumnos puntúan en un rango de notas más concentrado que el profesor (9-6 vs.10-4). Esto hace que las medias de las calificaciones de ambos grupos sean similares (7,8 profesor vs. 7,4 alumno). Los resultados obtenidos

corroboran, en lo que se refiere a la dispersión de notas otorgada por alumnos, los obtenidos por MacAlpine (1999) cuando mantiene que aunque la tarea evaluadora se tome en serio por los alumnos, sus expectativas de lo que es una buena exposición varían notablemente entre los alumnos evaluadores. Sin embargo, el hecho, de calcular la media de los 3 alumnos que califican contribuye a eliminar la dispersión al alza o a la baja concedida por alguno de ellos; de esta forma, aun en el caso de que algún estudiante individual resulte ser un juez poco riguroso, la fiabilidad de la nota media aumenta con el número de evaluadores.

Aunque en términos de los promedios obtenidos existe un relativo acuerdo en lo que se refiere a la correlación profesor-alumnos, las deficiencias detectadas en la evaluación de presentaciones deficientes con el objeto de ayudar al compañero debe traducirse en una mayor concienciación y formación del alumnado para conseguir un juicio justo que aumente la fiabilidad de la coevaluación académica entre pares. El estudio llevado a cabo detecta también los aspectos débiles en parámetros más subjetivos a evaluar o que requieren una mayor práctica por parte del alumno para su evaluación, como puede ser la organización de la información. En cualquier caso, se percibe que la participación del alumno en tareas de evaluación es claramente positiva, tanto como actividad de aprendizaje de lo que exige una buena presentación oral, como en el desarrollo del espíritu crítico del alumno a través de la evaluación. La adquisición de estas competencias es de clara utilidad para el futuro profesional de los estudiantes y conduce al aprendizaje para toda la vida (*life-long learning*).

Bibliografía

- Cheng, W. y Warren, M. (1999). Peer and teacher assessment of the oral and written tasks of a group project. *Assessment & Evaluation in Higher Education*, 24(3): 301–314.
- Durán, R., Sánchez-Reyes, S. y Beltrán, F. (2004). *La formación del profesorado de lengua inglesa en un contexto europeo*. Salamanca: Almar.
- Edwards Schachter, M. y Llopis Castelló, R. (2004). Situación de las Universidades Politécnicas Europeas en el Nuevo Espacio Europeo de Educación Superior. *Anuario Latinoamericano de Química (ALDEQ)*, XVIII: 172-180.
- Freeman, M. (1995). Peer assessment of groups of group work. *Assessment & Evaluation in Higher Education*, 20: 289-299.
- Goodale, M. 1998. *Professional presentations*. Milton: Cambridge University Press.

- Kwan, K.P. y Leung, R. (1996). Tutor versus peer group assessment of student performance in a simulation training exercise. *Assessment & Evaluation in Higher Education*, 21(3): 205-213.
- MacAlpine, J.M.K. (1999). Improving and encouraging peer assessment of student presentations. *Assessment & Evaluation in Higher Education*, 24 (1): 15-25.
- Matthews, C. y Marino, J. (1990). *Professional interactions: Oral communication skills in science, technology and medicine*. London: Prentice Hall.
- Milton, J. (2005). *Self and peer assessment*.
<http://www.lts.rmit.edu.au/renewal/assess/faq2.htm#judging.RMITUniversity> [ref. 5 junio 2005].
- Orsmond, P., Merry, S. & Reiling, K. (1996). The importance of marking criteria in the use of peer assessment. *Assessment & Evaluation in Higher Education*, 21(3): 239-250.
- Stefani, L.A.J. (1994). Peer, self and tutor assessment. *Assessment & Evaluation in Higher Education*, 18: 221-233.
- Sherpa, C. (2004). *Assessing group work*.
<http://www.utdc.vuw.ac.nz/resources/guidelines/GroupWork.pdf>. [ref. 5 junio 2005].

Apéndice A. Evaluación del proceso

Evaluación del proceso: 1 punto a repartir entre los miembros del grupo

Escala de aplicación y criterios:

- 3 = contribución muy alta
- 2 = contribución por encima de la media
- 1 = contribución por debajo de la media
- 0 = no contribuye
- -1 = contribución con carácter negativo.

Evaluación otorgada a	Carlos			Pilar			Merche		
Alumnos evaluadores	Carlos	Pilar	Merche	Pilar	Merche	Carlos	Merche	Pilar	Carlos
Ideas	3	3	2	2	2	1	1	2	2
Producción material	2	-1	0	3	2	2	0	1	2
Redacción informe	1	2	1	3	3	2	1	1	2
<i>Total</i>	<i>13</i>			<i>20</i>			<i>12</i>		

(Basado en Sherpa, 2004)

Puntuación media: $(13 + 20 + 12) = 45/3 = 15$

Reparto de 1 punto entre los 3

Puntuación obtenida por Carlos: $13/45 = 0.28$ p

Puntuación Pilar: $20/45 = 0.44$ p

Puntuación Merche: $12/45 = 0.26$ p.

Apéndice B. Evaluación de la presentación oral

Alumno evaluado:

Fecha:

Evalúa la presentación de tu compañero con una nota del 0 al 6 en cada uno de los apartados siguientes teniendo en cuenta los criterios que comprenden

CRITERIOS PARA EVALUACIÓN PRESENTACIÓN ORAL	
Exposición	<ul style="list-style-type: none">- Volumen- Contacto visual con la audiencia- Naturalidad- Velocidad de habla- Posturas y movimientos
Contenido	<ul style="list-style-type: none">- Claridad en la idea principal- Tema adecuado para el tiempo disponible- Tema adecuado para la audiencia- Tema desarrollado con detalles, hechos y ejemplos importantes- Tratamiento del tema en el tiempo asignado
Organización	<ul style="list-style-type: none">- Introducción- Transiciones- Puntos principales claramente resaltados- Desarrollo lógico de ideas- Conclusión
Lengua	<ul style="list-style-type: none">- Precisión de la comunicación- Vocabulario apropiado para la audiencia- Pronunciación/entonación- Variedad de expresiones

Guía para la coevaluación del trabajo en equipo

Fernando González Ladrón de Guevara

En este estudio se utiliza metodología cualitativa y cuantitativa con el fin de elaborar una guía que permita aplicar un protocolo para evaluar el trabajo en equipo, tanto por sus pares (otros equipos) como por el profesor. Este estudio trata también de validar la fiabilidad de dicha guía. El trabajo de campo se ha llevado a cabo en la asignatura Ingeniería de Producción y Gestión de Operaciones, impartida en la Escuela Técnica Superior de Ingenieros de Telecomunicación, durante el curso 2004-2005. Los resultados obtenidos muestran la relación entre la evaluación realizada por equipos y por el profesor, al considerar los diferentes aspectos incluidos en una guía para la coevaluación del trabajo en equipo.

Palabras clave: *coevaluación, criterios de evaluación, evaluación entre pares, evaluación del proceso y del producto, trabajo en equipo, fiabilidad.*

1. Introducción

La evaluación y el control se emplean para determinar el grado de asimilación de los conocimientos impartidos, aptitudes, motivación y actitudes adquiridas por el discente y para comprobar y mejorar el proceso de enseñanza-aprendizaje como sistema total (Zabalza, 2000). Entre los objetivos específicos están:

- Conocer el grado de cumplimiento de los objetivos y valorar el rendimiento del discente. Como plantea Rotger Amengual (1979), la medida, para que tenga verdadero significado, precisa ser evaluada e interpretada en relación a una norma ya establecida, para obtener información que indique en qué medida se han alcanzado los objetivos.

- Diagnosticar dificultades de aprendizaje: tener la mayor información posible de partida para analizar las causas que puedan ocasionar las deficiencias en las metas propuestas, métodos y enfoques, y actuar en consecuencia.
- Mejorar los planes y las técnicas de trabajo y procurar evitar la repetición de errores, reforzando oportunamente aquellas áreas de estudio en que el aprendizaje es insuficiente.
- Considerar la conveniencia y viabilidad de los programas a la vista de los resultados.

De estos objetivos se deduce que evaluar no es castigar o premiar con puntuaciones, sino un permanente control de calidad que incide en la mejora del producto.

Los alumnos consideran la evaluación, en la mayoría de los casos, como una "trampa" que hay que superar. Esta sensación se refuerza por el desconocimiento de los criterios de evaluación que va a utilizar el profesor o incluso de algún ejemplo de examen de años anteriores. La evaluación, basada únicamente en la realización de una prueba escrita, refuerza la actitud de estudiar para aprobar y, por ende, conduce a un aprendizaje superficial, poco duradero e insatisfactorio, en definitiva poco significativo, que acarrea la frustración del alumnado y del profesor.

La nueva política de educación superior implica una profunda transformación en entornos universitarios, donde la educación centrada en la enseñanza pasa a la educación centrada en el aprendizaje. Es necesario un cambio de actitud del docente y del alumno para reorientar la evaluación de forma que satisfaga los requerimientos demandados (Baldwin, 2005).

La coevaluación o evaluación entre pares presenta unas características que la convierten en alternativa a la evaluación tradicional, especialmente en su vertiente sumativa. Permite a los alumnos reflexionar de forma crítica sobre los resultados obtenidos. La reflexión sobre el trabajo ajeno, también con una componente de autocrítica, es una habilidad necesaria no sólo en el mundo académico sino en el mundo laboral y, como apunta Pfeffer (1994), constituye un objetivo digno de ser tenido en cuenta. La concienciación del alumnado sobre dicha utilidad y planteado como un proceso convenientemente estructurado, permite implicar al alumno y aumentar su satisfacción (Freeman, 1995). Es evidente que precisa de un cierto adiestramiento y también de una realimentación: recibir información comparativa de los resultados obtenidos por sus pares.

Una de las debilidades de la coevaluación es su fiabilidad. Según Kwan y Leung (1996), ésta plantea dudas sobre su conveniencia tanto al alumno como al profesor. Por ello, disponer de unos criterios detallados de evaluación contribuye a mejorar la estructuración del proceso y aumenta su calidad.

El presente estudio se lleva a cabo en la asignatura *Ingeniería de Producción y Gestión de Operaciones*, asignatura optativa que se imparte en la Escuela Técnica Superior de Ingenieros de Telecomunicación de la Universidad Politécnica de Valencia (UPV). En el estudio intervienen dos grupos con 55 y 25 alumnos respectivamente. El grupo menos numeroso realiza el proceso de enseñanza-aprendizaje en lengua inglesa. La metodología que se utiliza es el enfoque de Aprendizaje Basado en Proyectos. Las tareas se hacen en equipo, cada uno formado por un número que oscila entre 4-6 personas que desarrollan un proyecto. Los alumnos pueden elegir entre dos opciones a la hora de realizar dicho trabajo:

- a) Aplicar una lista de cuestiones a una empresa real para definir su sistema de operaciones. Este trabajo responde a preguntas de aplicación práctica de cada bloque temático. El ámbito es una empresa del sector industrial o de servicios que cada equipo escoge.
- b) Trabajar en profundidad sobre modelos cuantitativos. Definición de un caso práctico con herramienta de autor multimedia.

Los resultados que cada equipo obtiene se recogen en una página web de diseño propio que incorpora además las actividades de aula, así como indicadores del proceso de trabajo, como el diario de grupo o las conclusiones parciales obtenidas, entre otros. Permite, además, que cada equipo de trabajo pueda conocer y analizar los resultados logrados por sus compañeros, al promoverse y valorarse el análisis continuo de dicha página.

Las páginas web están alojadas en un servidor Linux y cada grupo recibe una zona de desarrollo para su página web. Además se utiliza exhaustivamente la microweb de la asignatura para poner a disposición de los alumnos los recursos didácticos y como foro de comunicación entre alumnos y profesor.

En esta asignatura se ha participado en diferentes proyectos de innovación. Desde el año 2001 hasta el 2003, se han trabajado estrategias para la mejora del proceso de enseñanza-aprendizaje y, específicamente, los métodos de evaluación. Durante el curso 2004-2005 el método de evaluación es de tipo cooperativo, en el que participan los alumnos, agrupados en equipos de trabajo y el profesor, con una ponderación del 50% en cada caso, según queda reflejado en la tabla 1.

<i>Ítems</i>	<i>Evaluación por profesor</i>	<i>Evaluación por pares</i>
Prácticas de aula	50%	50%
Proyecto	50%	50%

Tabla 1. Ponderación en la evaluación cooperativa

Los alumnos utilizan una plantilla de criterios (ver Apéndice) para valorar aspectos del proceso y del resultado, cuya validez y mejora constituye uno de los objetivos de este trabajo. Se exige una asistencia a clase del 80% para poder ser evaluado de esta forma, de lo contrario existe la alternativa de realizar una prueba escrita al final de curso.

Cada equipo, además de mostrar los resultados de su trabajo en la correspondiente página web, realiza diferentes presentaciones orales sobre características parciales de su proyecto. Estas presentaciones se completan con una presentación final, en la que participan todos los componentes del equipo que trabajan en un mismo proyecto, con una duración de 30', seguida de un turno de ruegos y preguntas.

La última actividad consiste en cumplimentar la parrilla de evaluación y hacerla llegar al profesor por medio de correo electrónico. Tras el oportuno registro y proceso, las calificaciones finales se hacen públicas.

1.1. Objetivos

El estudio tiene como objetivo el diseño y mejora de una guía para evaluar el trabajo en equipo, tanto por sus iguales (otros equipos) como por el profesor. Para ello se describe el proceso de evaluación en su conjunto, de forma que se puedan detectar debilidades y elaborar estrategias que aumenten la calidad, la fiabilidad y la contribución al cumplimiento de los objetivos.

2. Materiales y método

El proceso de evaluación compartida se presenta a los alumnos como una oportunidad para reflexionar sobre el trabajo de sus compañeros y adquirir la competencia de valoración y análisis crítico. Se comentan también aspectos éticos y de precisión de dicho proceso.

Los alumnos, a lo largo del curso, cuentan con los trabajos realizados por equipos de años anteriores disponibles en la microweb de la asignatura. Además, se comentan aquellos resultados de calidad para ilustrar los contenidos teóricos. Esta actividad permite a los alumnos apreciar los resultados que se pretende alcanzar y cómo lo han logrado otros equipos. De esta forma, la microweb de la asignatura es una especie de portafolio docente que recoge los trabajos de los distintos equipos de alumnos, tanto del curso presente como de cursos anteriores.

Las actividades se realizan desde principio de curso por los equipos formados inicialmente. De esta forma, el trabajo es continuo a lo largo de todo el curso, y los

diferentes equipos contrastan la evolución y resultados de sus compañeros. Cada equipo realiza un *blog* que recoge las opiniones y sugerencias de otros equipos visitantes de su página web.

Los datos relativos a la evaluación por equipos se reciben por correo electrónico y se almacenan en una base de datos *Access*, diseñada *ad hoc* para su posterior tratamiento estadístico con el programa informático SPSS 12.0.

2.1. Guía para el proceso de evaluación

La guía para la evaluación de los equipos (ver Apéndice) se resume en la tabla 2.

<i>Ítem</i>	<i>Explicación</i>	<i>% sobre puntuación final</i>
Interface	Ergonomía. Actualización. Innovación.	25%
Contenidos	Completa. Han realizado todas las tareas. Corrección. Relaciones entre contenidos y proyecto.	25%
Valor añadido	Enlaces interesantes. Nuevas ideas. Utilidad. Un modelo a seguir.	25%
Capacidad comunicación	Interesante, dinámica, coherente, ajustada a tiempo.	15%
Imagen equipo	Coordinación, reparto de tareas.	10%

Tabla 2. Guía para la evaluación de los equipos

Se pretende que esta herramienta sea un protocolo de actuación, de manera que ayude en el proceso de reflexión al evaluar el trabajo desempeñado por los compañeros. En este protocolo se valoran los siguientes aspectos.

1. *Interface*. Los equipos desarrollan una página web donde, a lo largo del curso, aparece el desarrollo y progreso de las distintas actividades: *blog* de reuniones, calendario, problemas, resolución de tests de autoaprendizaje, análisis de casos y el desarrollo de proyectos. Se pretende valorar la ergonomía de dicha página web: claridad, organización, facilidad de uso y consulta. De igual modo, se valora cómo se va actualizando a lo largo del curso, si se

incorporan las distintas actualizaciones puntualmente o si la hoja web ha estado “dormida” durante mucho tiempo. Finalmente, se considera también la componente de innovación de dicha página. El porcentaje de este apartado supone un 25% de la calificación total.

2. *Contenidos técnicos.* Este apartado recoge la valoración respecto a la consecución o no de las tareas, el planteamiento de los ejercicios, la relación entre diferentes contenidos conceptuales, la aportación de forma más exhaustiva de aspectos que han sido presentados en clase de manera superficial y la innovación del método de resolución de los problemas utilizado, de manera que puedan ser un modelo a seguir. Otro de los valores a considerar son los enlaces interesantes a otras páginas web que aportan documentación adecuada y amplían los contenidos expuestos. En definitiva, se puede decir si los contenidos conceptuales presentes en la página web y la presentación son brillantes. También se tiene en cuenta las respuestas que se facilitan en el turno de aclaraciones tras la exposición oral. La valoración de este apartado aporta el mismo porcentaje que el anterior, el 25% de la calificación final.
3. *Capacidad de comunicación.* Los conceptos considerados son los siguientes: el interés de la presentación del trabajo, si la presentación es dinámica, equilibrada y ajustada al tiempo fijado y la participación de todos los alumnos del grupo (este aspecto tiene especial interés en el grupo que utiliza el inglés como lengua vehicular). El peso de este apartado supone el 15% respecto de la calificación final.
4. *Imagen de equipo.* Aquí se evalúa el proceso al considerar las siguientes evidencias: las presentaciones orales que el equipo realiza, el *blog* del equipo con las anotaciones del cuaderno de reuniones y la coordinación de tareas. En definitiva, se evalúa el funcionamiento y coordinación del equipo, es decir, si todos los componentes del grupo han trabajado en equipo. El porcentaje de este apartado es del 10% respecto al total.

En la guía existe un apartado que recoge las observaciones de los alumnos, indicaciones cualitativas que muestran aspectos que los apartados anteriores de la guía no aportan; se trata de matizaciones abiertas que completan las demás valoraciones y pueden contribuir a mejorar la guía.

3. Resultados

En primer lugar, se realiza un análisis descriptivo de los datos obtenidos en el trabajo de campo con el grupo de castellano y, posteriormente, se consideran los datos relativos al grupo con aprendizaje en lengua inglesa. Por último, se explora la relación entre los cinco aspectos con los datos obtenidos del grupo en castellano.

3.1. Grupo en castellano

Se dispone de 110 series de calificaciones procedentes de las valoraciones realizadas por cada uno de los 11 equipos sobre los 10 restantes (cada equipo califica al resto). Cada serie incorpora la valoración de los cinco ítems correspondientes a los diferentes aspectos de la parrilla de evaluación, con un total de 550 valores. En la tabla 3 se muestra la media aritmética de las diez calificaciones obtenidas por cada uno de los equipos, la aportada por el profesor y la diferencia entre ambos valores.

<i>Calificaciones proporcionadas por los equipos</i>	<i>Calificaciones proporcionadas por el profesor</i>	<i>Diferencia</i>
7,9	7,0	0,97
7,6	7,0	0,63
8,8	9,1	0,34
8,0	7,6	0,39
8,3	8,3	0,04
8,9	9,6	0,73
8,4	7,8	0,56
7,7	6,3	1,45
8,4	8,9	0,52
7,7	6,6	1,12
8,1	7,7	0,41

Tabla 3. Medias de las calificaciones de los equipos de trabajo otorgadas por el resto de los equipos y por el profesor

Se observa que la discrepancia no es elevada y que la media aritmética de las diferencias tiene un valor de 0,65. La tabla 4 muestra los estadísticos de las variables anteriores.

	<i>N</i>	<i>Media</i>	<i>Desviación típica</i>	<i>Máximo</i>	<i>Mínimo</i>	<i>Rango intercuartílico</i>
Calificaciones profesor	11	7,81	1,08	9,6	6,26	1,9
Calificaciones equipos	11	8,16	0,42	8,9	7,63	0,7

Tabla 4. Estadísticos centrales y de dispersión de las calificaciones aportadas por el profesor y los equipos

Siguiendo a Kwan y Leung (1996), la media de las notas de los equipos (8,17) queda dentro del intervalo formado por la media de las notas del profesor, como medida central, con +/- una desviación típica de amplitud, [6,73; 8,89]. Estos resultados informan sobre el acuerdo entre las evaluaciones de los alumnos y las del profesor, puesto que la media de las calificaciones de los equipos se encuentra entre el intervalo presentado. El rango es de 3,37 para la variable calificaciones del profesor y 1,27 para las calificaciones de equipos. La figura 1 muestra las diferencias entre las calificaciones del profesor y la media de los equipos tomando como factor cada uno de ellos.

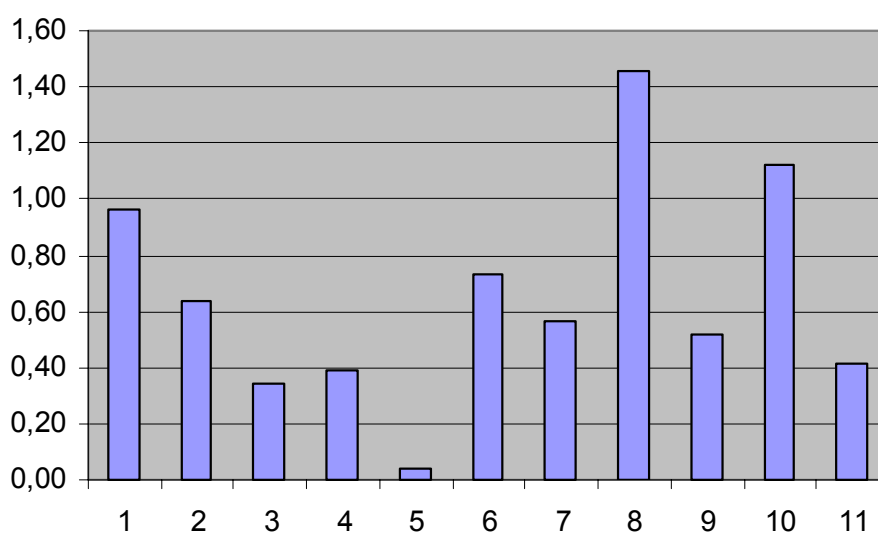


Figura 1. Diferencias entre las calificaciones del profesor y la media de los equipos

Como se aprecia en la figura 2, el rango intercuartílico de las calificaciones del profesor es más amplio que el resultante de considerar las calificaciones de los equipos. Se observa una mayor desviación en las notas del profesor, quien discrimina más.

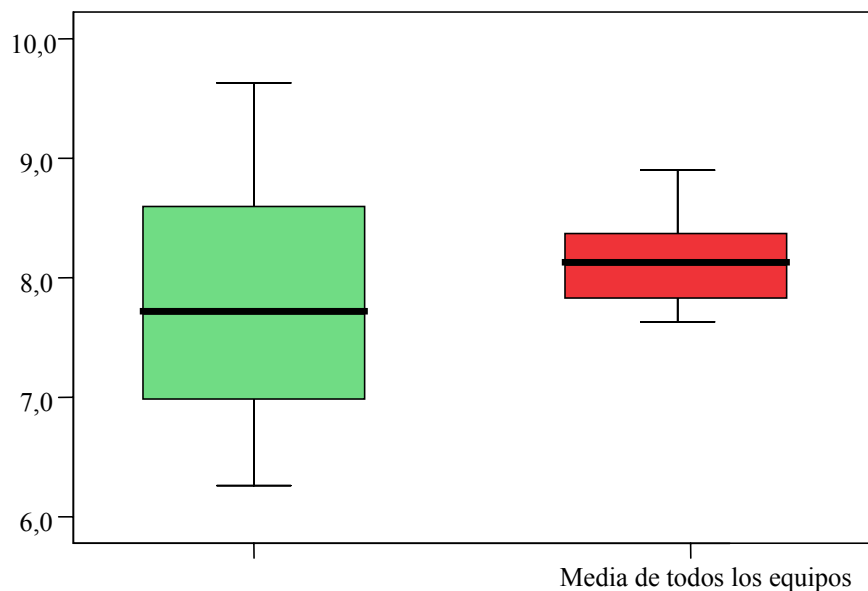


Figura 2. Calificaciones del profesor y de los equipos

3.2. Grupo en inglés

Los valores obtenidos de las calificaciones otorgadas por el profesor y de las medias aritméticas recibidas por los distintos equipos (5 casos) se muestran en la tabla 5.

<i>Calificaciones proporcionadas por los equipos</i>	<i>Calificaciones proporcionadas por el profesor</i>	<i>Diferencia</i>
8,8	9,3	0,53
8,3	8,3	0,06
9,0	10,0	0,98
7,6	6,5	1,11
8,1	7,7	0,41

Tabla 5. Medias de las calificaciones de los equipos de trabajo otorgadas por el resto de los equipos y por el profesor al grupo en inglés

Los resultados de este grupo son mejores que los de castellano con un valor medio de 8,36 de la variable notas del profesor. El valor de la media aritmética de las diferencias es similar al grupo de castellano y tiene un valor de 0,67. La tabla 6 muestra los estadísticos descriptivos.

	<i>N</i>	<i>Media</i>	<i>Desviación típica</i>	<i>Máximo</i>	<i>Mínimo</i>	<i>Rango intercuartílico</i>
Calificaciones profesor	5	8,36	1,36	10	6,53	2,55
Calificaciones equipos	5	8,37	0,55	9,03	7,63	1,04

Tabla 6. Estadísticos centrales y de dispersión de las calificaciones aportadas por el profesor y los equipos en el grupo en inglés

El intervalo formado por la media de las notas del profesor +/- la desviación típica es [7; 9,72]. En este caso, la media aritmética de las calificaciones de los alumnos está situada también dentro del intervalo, diferenciándose tan sólo una centésima de la media de las notas del profesor. Los alumnos del grupo en lengua inglesa tienen una mayor variabilidad en sus calificaciones que los del grupo en castellano, siendo el rango entre el valor máximo y mínimo de 3,47 para las calificaciones del profesor y 1,4 para las calificaciones de los alumnos. La figura 3 muestra el gráfico de caja que corrobora las afirmaciones anteriores.

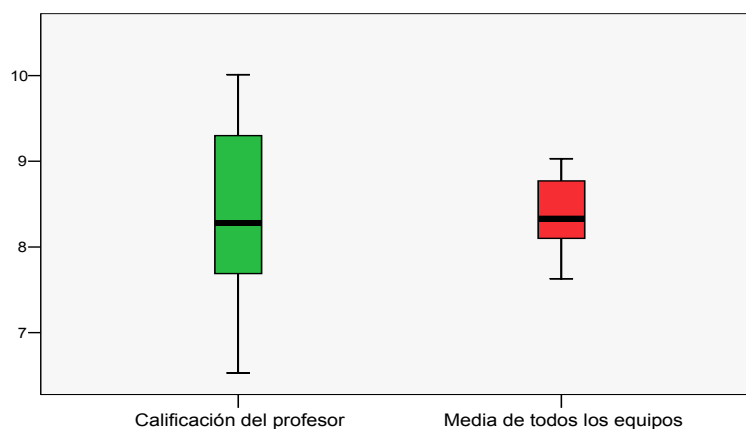


Figura 3. Calificaciones del profesor y de los equipos en el grupo en inglés

En ambos casos se aprecia una menor variabilidad en las calificaciones de los alumnos, los alumnos tienden a dar valores similares discriminando menos, no conceden valores altos y penalizan menos a los peores equipos. Estos resultados concuerdan con los obtenidos por Freeman (1995).

A continuación se analizan, de forma exploratoria, las variables indicativas de cada uno de los aspectos a considerar en la parrilla de evaluación, para el grupo en castellano que es el que tiene un mayor número de observaciones.

3.3. Análisis por variables

Se dispone de 110 observaciones correspondientes a las valoraciones de cada uno de los equipos sobre el resto de equipos. En la tabla 7 se muestran los estadísticos básicos de la variable Total, suma de las valoraciones de los ítems de la plantilla.

	<i>N</i>	<i>Media</i>	<i>Desviación típica</i>	<i>Máximo</i>	<i>Mínimo</i>	<i>Rango intercuartílico</i>
Total	110	8,16	0,87	9,8	5,45	1,12

Tabla 7. Estadísticos de la variable Total

El gráfico de caja de la figura 4 muestra la existencia de dos valores atípicos. Corresponden a valoraciones de uno de los equipos (equipo número 6); se trata de los casos 65 y 68 (que toman los valores 5,45 y 5,95 respectivamente). El test de Kolmogorov-Smirnov nos da un valor de 0,104 que garantiza que los datos se ajustan a una distribución normal.

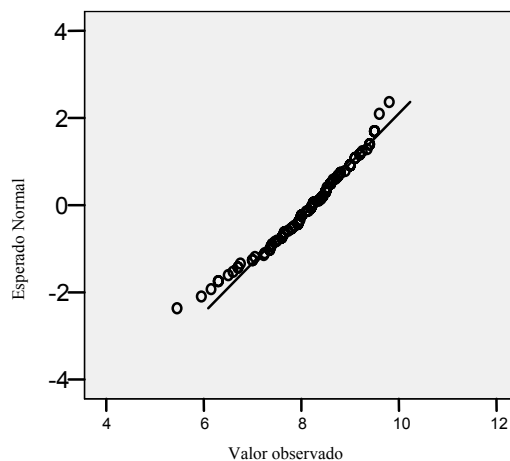


Figura 4. Gráfico de probabilidad normal Q-Q

A continuación se presentan los estadísticos descriptivos de las variables que reflejan los resultados de la parrilla de evaluación (ver Tabla 8).

	<i>N</i>	<i>Media</i>	<i>Desviación típica</i>	<i>Máximo</i>	<i>Mínimo</i>	<i>Rango intercuartílico</i>
Interface	110	8,22	1,16	10	6	1,56
Contenido técnico	110	8,43	1,10	10	4	1
Valor añadido	110	7,40	1,77	10	2	1,7
Capacidad comunicación	110	8,24	1,23	10	4,7	1,75
Imagen de equipo	110	9	1,00	10	6	2
Total	110	8,16	0,87	9,8	5,45	1,12

Tabla 8. Estadísticos descriptivos de los ítems de la parrilla de evaluación

La variable que presenta un valor menor de su media aritmética es *Valor Añadido*. Este hecho es lógico al tener un planteamiento de excelencia en su definición. El resto de variables tienen un comportamiento similar.

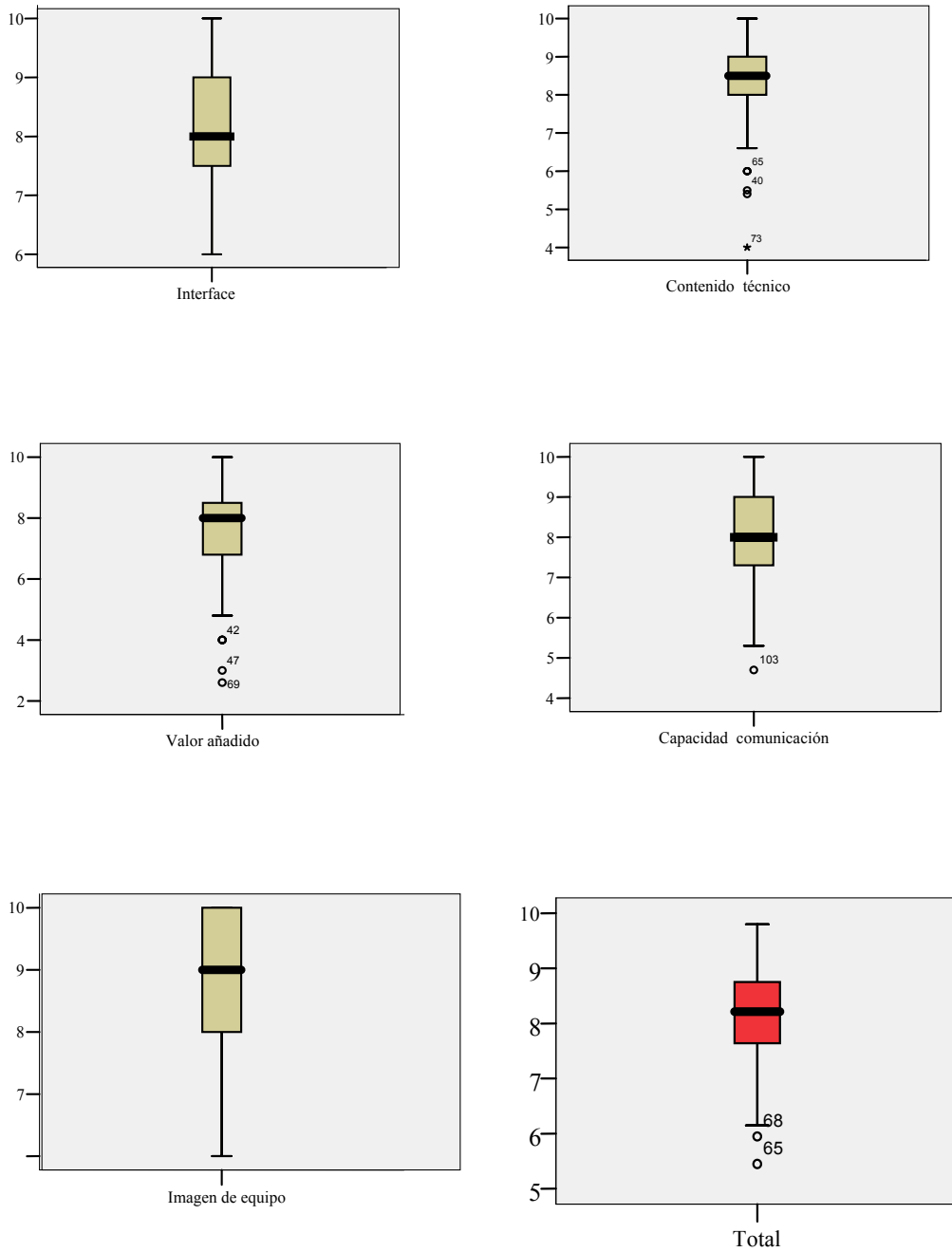


Figura 5. Diagrama de caja de los ítems de la parrilla de evaluación

La variable *Interface* no posee valores atípicos (ver Figura 5) y no se ajusta a una distribución normal (no supera el test de Kolmogorov-Smirnov).

Contenido técnico tiene un factor de curtosis de 2,037 que descarta su normalidad (corroborado por el test de Kolmogorov-Smirnov); según muestra la figura 5, presenta valores atípicos y un valor extremo inferior con valor 4 que se aparta una distancia superior a 1,5 cuartiles de la media. Se trata de la variable con la menor dispersión y concentrada en la parte superior de la escala. Es necesario pues redefinir los criterios de forma que se pueda apreciar con claridad las condiciones para discriminar los resultados.

Valor añadido no supera el test de ajuste a la distribución normal de Kolmogorov-Smirnov. Es la variable que presenta mayor número de valores atípicos en la gama baja de su distribución, tal y como queda de manifiesto en el diagrama de caja correspondiente (ver Figura 5). Presenta un valor de la media aritmética inferior a las variables anteriores con un grado de dispersión mayor. Se trata pues de una variable (como podía preverse de su propia definición) que incorpora una posibilidad mayor de discriminación.

La variable *Capacidad de comunicación* no supera el test de normalidad, con 21 casos que han obtenido la máxima calificación (10). Como se observa en la figura 5, presenta un único valor atípico. Se trata de una variable en la que quizá los alumnos han discriminado menos y ha habido una tendencia generalizada a otorgar la máxima calificación.

Imagen de equipo alcanza la media aritmética más elevada (con un valor para la mediana de 9, siendo la moda 10 con un valor de 45). En este caso es más difícil discriminar y aportar valores inferiores. No supera el test de normalidad. Presenta un único caso atípico que coincide con el mínimo. La capacidad para discriminar es muy pequeña. Su diagrama de caja, según aparece en la figura 5, confirma las afirmaciones anteriores al no tener rama superior.

Si se considera la variable *Total*, atendiendo al equipo de trabajo que evalúa y se analizan las calificaciones proporcionadas por cada uno de los once equipos del grupo en castellano, se obtienen los estadísticos descriptivos que aparecen en la tabla 9.

<i>Evaluador</i>	<i>Media</i>	<i>Desviación Típica</i>	<i>Máx.</i>	<i>Mín.</i>	<i>Rango intercuartílico</i>
1	7,46	0,53	8,20	6,60	0,82
2	8,11	0,46	8,65	7,35	0,70
3	8,39	0,49	9,40	7,85	0,70
4	7,77	1,21	9,40	6,30	2,35
5	8,60	0,45	9,35	7,65	0,41
6	6,88	0,85	7,95	5,45	1,51
7	8,90	0,60	9,60	8,00	1,08
8	8,47	0,85	9,50	7,00	1,40
9	8,74	0,51	9,80	7,80	0,45
10	8,03	0,41	8,75	7,60	0,74
11	8,42	0,81	9,50	7,00	1,23

Tabla 9. Estadísticos descriptivos de la variable Total considerada por equipos de trabajo

Puede observarse que el equipo que ha emitido calificaciones más elevadas, por término medio, alcanza un valor de 8,9 (equipo 7) y el que tiene un valor medio menor toma un valor de 6,88 (equipo 6). Así pues, se percibe una diferencia de 2,02 puntos en las calificaciones medias; este resultado, unido a los valores que toma la desviación típica, permite afirmar la diferencia de orientación “a la baja” en las calificaciones del equipo 6 (valor máximo de 7,95 y mínimo de 5,45). El resto de los equipos presenta un comportamiento similar en sus estadísticos de centralización (ver Figura 6). Si se consideran las desviaciones típicas puede observarse que son similares para todos los equipos, excepto para el equipo 4 que tiene un valor de 1,21 y discrimina mucho más (valor máximo 9,4 y valor mínimo 6,3).

La realización de un análisis de varianza sobre la variable *Total* con el factor de grupo evaluador no da ningún resultado coherente porque la prueba de homogeneidad de varianzas (estadístico de Levene) no presenta un p-valor mayor que 0,05; es decir, no hay igualdad de varianzas. Este hecho queda también de manifiesto al considerar los gráficos de caja (ver Figura 6) en el que se presenta la variable total tomando como factor de grupo el equipo evaluador.

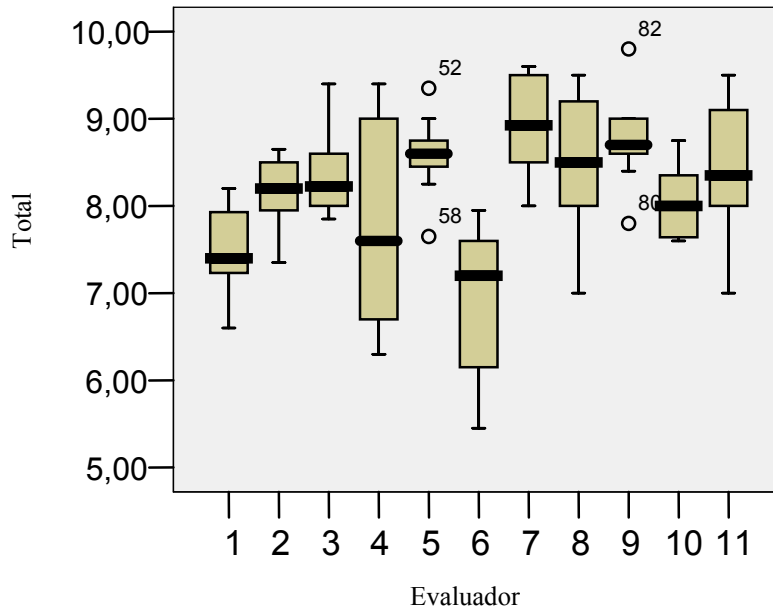


Figura 6. Variable Total respecto al grupo evaluador

Tras analizar las correlaciones bivariadas entre las cinco variables que forman parte de la parrilla evaluadora y comprobar que los valores no pueden suponerse que pertenezcan a una distribución normal, se ha utilizado la Rho de Spearman y la Tau-b de Kendall. Los valores obtenidos se muestran en la tabla 10.

Correlaciones

		Interface	Contenido técnico	Valor añadido	Capacidad comunicac.	Imagen equipo
Interface	Coefficiente correlación	1,000	,280 **	,340 **	,144	,121
	Sig. (2-tailed)	.	,003	,000	,135	,207
	N	110	110	110	110	110
Contenido técnico	Coefficiente correlación	,280**	1,000	,409 **	,322 **	,161
	Sig. (2-tailed)	,003	.	,000	,001	,094
	N	110	110	110	110	110
Valor añadido	Coefficiente correlación	,340**	,409**	1,000	,201 *	,147
	Sig. (2-tailed)	,000	,000	.	,036	,126
	N	110	110	110	110	110
Capacidad comunicación	Coefficiente correlación	,144	,322**	,201*	1,000	,422
	Sig. (2-tailed)	,135	,001	,036	.	,000
	N	110	110	110	110	110
Imagen equipo	Coefficiente correlación	,121	,161	,147	,422**	1,000
	Sig. (2-tailed)	,207	,094	,126	,000	.
	N	110	110	110	110	110

** . Correlación significativa al nivel 0.01.

* . Correlación significativa al nivel 0.05.

Tabla 10. Correlaciones de los ítems de la parrilla de evaluación

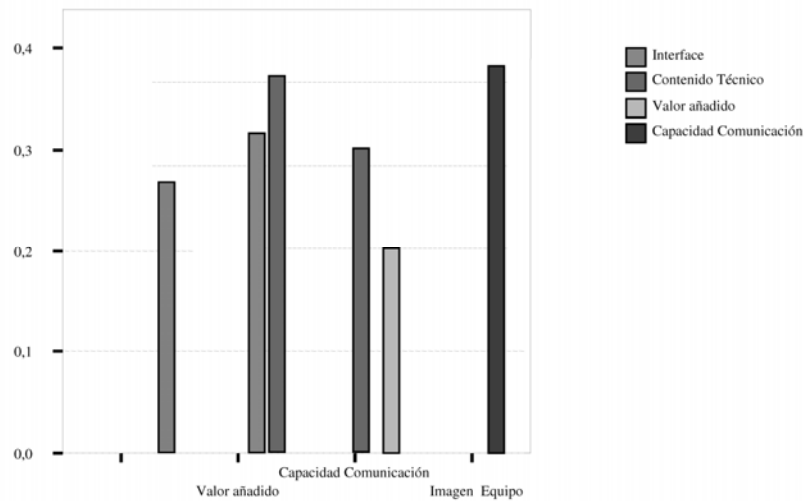


Figura 7. Correlaciones significativas entre los ítems de la parrilla de evaluación. Valores de Rho de Spearman

Como se muestra en la figura 7, las correlaciones significativas tienen un valor del coeficiente de Spearman relativamente bajo. Las más elevadas se dan entre *Imagen de equipo* y *Capacidad de comunicación* (0,422) y entre *Valor añadido* y *Contenidos técnicos* (0,409). Cabe destacar las correlaciones existentes entre *Valor añadido* y *Interface*, entre *Capacidad de comunicación* y *Contenidos técnicos* y finalmente entre *Interface* y *Contenidos técnicos*.

Los resultados obtenidos sugieren la posibilidad de cambiar la ponderación de la variable *Contenidos técnicos* para potenciar variables que midan la excelencia del trabajo como *Valor añadido*.

Por otra parte, se consideran la nota media recibida, nota media aportada (del conjunto de las valoraciones), desviación típica de dichas valoraciones y rango de las mismas (ver Tabla 11). Las cuatro variables superan el test de normalidad de Kolmogorov-Smirnov.

<i>Nº Grupo</i>	<i>Nota media recibida</i>	<i>Nota media aportada</i>	<i>Desviación valoraciones</i>
1	7,9	7,46	0,53
2	7,6	8,11	0,46
3	8,8	8,39	0,49
4	8,0	7,77	1,21
5	8,3	8,60	0,45
6	8,9	6,88	0,85
7	8,4	8,90	0,60
8	7,7	8,47	0,85
9	8,4	8,74	0,51
10	7,7	8,03	0,41
11	8,1	8,42	0,81

Tabla 11. Nota media recibida, aportada y desviación de valoraciones por los distintos grupos

El análisis de las correlaciones bivariadas permite afirmar que no hay correlaciones significativas entre las variables.

4. Conclusiones

En la experiencia de evaluación participativa que aquí se presenta, los alumnos se evalúan entre sí y el profesor interviene y controla la calificación final. Se constata la utilidad de la coevaluación como alternativa a una evaluación más tradicional y su grado de acuerdo con la evaluación realizada por el profesor. Se observa que la concordancia se ve mejorada, especialmente, cuando se realiza la concienciación sobre la utilidad de la experiencia de evaluar, se estructura el proceso de coevaluación y se ofrece garantía al alumnado sobre la fiabilidad. No obstante, se ha planteado también a los alumnos que las irregularidades en el proceso podrían provocar que se abandonara este método de evaluación.

Para conseguir sistematización en la evaluación es útil disponer de una guía que contraste los distintos rasgos de calidad de los trabajos de los alumnos. La guía de evaluación tiene en cuenta los aspectos *Interface*, *Contenidos Técnicos*, *Valor Añadido*, *Capacidad de Comunicación* e *Imagen de Equipo*. La fiabilidad de la evaluación aumenta cuando aumenta el número de aspectos valorados, al intervenir un número mayor de evaluadores y al trabajar con valores medios. Según los resultados de este estudio, deben primarse aquellos aspectos que reflejan la excelencia en el trabajo y que los alumnos son capaces de discriminar mejor, como es el caso de la variable *Valor Añadido*, sobre aspectos como *Capacidad de Comunicación*, que es un objetivo transversal y que se puede ver reflejado en la variable *Imagen de Equipo*. Se detectan aspectos que deben incorporar una descripción detallada de los requerimientos necesarios para conseguir una calificación máxima, como es el caso de la variable *Contenidos*. En general, los valores de las calificaciones medias aportadas por los diferentes equipos reflejan una orientación similar, tanto al considerar sus valores centrales como la dispersión de los valores.

En definitiva, se percibe un aumento de la motivación del alumnado, dado que el conocer que el propio trabajo está siendo analizado por los compañeros fomenta la implicación del alumno y, en consecuencia, la calidad del resultado.

Como futuras líneas de trabajo se pretende incorporar la evaluación del proceso mediante la evaluación intragrupo, es decir, que cada uno de los componentes que integran un equipo de trabajo evalúe a sus iguales. Una posible estrategia sería otorgar al grupo una puntuación global, obtenida de las puntuaciones obtenidas por cada componente del equipo y sugerir a éstos que se repartan entre ellos los puntos obtenidos. Para tal fin, sería interesante definir unas parrillas de criterios consensuados que ayuden a facilitar dicho reparto y que valoren aspectos tales como contribución individual, ideas aportadas, implicación y responsabilidad para finalizar los trabajos, participación en la presentación del trabajo, etc.

Finalmente, el incorporar al proceso de evaluación la negociación, buscando el consenso en la definición de los ítems de la parrilla y el porcentaje específico de cada uno de ellos respecto de la calificación total, puede aportar información significativa. De igual modo, aunque no se han recogido comentarios en la sección de observaciones relativos a la falta de claridad en cada ítem o sugiriendo modificaciones de los porcentajes, la elaboración de ítems más desglosados y la incorporación de elementos de valoración estándar para cada una de las características a evaluar podría aumentar de forma significativa la calidad del proceso de evaluación.

Bibliografía

- Baldwin, G. (2005). *Getting started with assessment. Basic advice for people new to university teaching*.
www.cshe.unimelb.edu.au/assessinglearning [ref. 20 noviembre 2005].
- Freeman, M. (1995). Peer assessment of groups of group work. *Assessment & Evaluation in Higher Education*, 20: 289-299.
- Kwan, K. P. y Leung, R. (1996). Tutor versus peer group assessment of student performance in a simulation training exercise. *Assessment & Evaluation in Higher Education*, 21(3): 205-213.
- Magin, D. J. (2001). A novel technique for comparing the reliability of multiple peer assessments with that of single teacher assessments of group process work. *Assessment & Evaluation in Higher Education*, 26(1): 53-63.
- Martínez de Sousa, J. (2003). *Manual de estilo de la lengua española* (2ª ed.). Gijón: Ediciones Trea.
- Orsmond, P., Merry, S. y Reiling, K. (1996). The importance of marking criteria in the use of peer assessment. *Assessment & Evaluation in Higher Education*, 21(3): 239-250.
- Pérez, C. (2005). *Técnicas estadísticas con SPSS 12*. Madrid: Pearson.
- Pfeffer, J. (1994). *Competitive advantage through people: Unleashing the power of the work force*. Boston: Harvard Business School Press.
- Rotger Amengual, B. (1979). *El proceso programador en la escuela*. (3ª ed.). Madrid: Editorial Escuela Española.
- Zabalza, M. A. (2000). *Diseño y desarrollo curricular*. (8ª ed.). Madrid: Narcea Ediciones.

Apéndice. Guía para la evaluación de los equipos

This table is intended to help evaluate each team and its work, both work done in class and on the project.

La presente Matriz pretende facilitar la evaluación de cada equipo y su trabajo, tanto el realizado en clase como el trabajo de proyecto.

Assessment values range from 1 to 10 and each factor has its specific weight, which is as follows:

La puntuación varía de 1 a 10 y cada factor tiene su peso específico. Estos son los siguientes:

<i>Item</i>	<i>Explanation / Aspectos a evaluar</i>	<i>%</i>
Interface	<i>Web page interface. Clear & easy to understand. Updated on time. Innovation/Interface de la página web. Clara y fácil de entender. Actualizada. Innovadora.</i>	2,5
Technical Contents/Contenidos técnicos	<i>All tasks have been done. Correct. Comprehensive. Good relationship between contents & project firm./Se han hecho todas las tareas. Correcta. Completa. Buenas relaciones entre contenidos y la empresa de proyecto.</i>	2,5
Added Value/Valor añadido	<i>Useful links. New ideas. It has helped us. A model to follow. Brilliant./Enlaces útiles. Nuevas ideas. Nos ha ayudado. Un modelo a seguir. Brillante</i>	2,5
Communication Ability/Capacidad de comunicación	<i>Presentation: Interesting, dynamic, coherent./Presentación: interesante, dinámica, coherente.</i>	1,5
Team Image/Imagen de equipo	<i>All the members of the group seem to have worked together./Parece que todos los componentes del equipo han trabajado juntos.</i>	1,0

TOTAL: 10 POINTS/PUNTOS.

This matrix is only a guide to facilitate your evaluation.

Esta matriz es sólo una guía para facilitar vuestra evaluación.

La evaluación compartida: investigación multidisciplinar

Equipo/*Team*: Fecha/*Date*:

Proyecto/*Project*:

<i>Item</i>	<i>Weight/ Porcentaje</i>	<i>Score/ Puntuacion</i>
Interface	2,5	
<i>Technical Contents/Contenidos técnicos</i>	2,5	
<i>Added Value/Valor añadido</i>	2,5	
<i>Communication Ability/Capacidad de comunicación</i>	1,5	
<i>Team Image/Imagen de equipo</i>	1,0	

TOTAL VALUE/ VALOR TOTAL:

REMARKS/OBSERVACIONES:

.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

Evaluación, coevaluación y autoevaluación del trabajo en grupo en la lectura de mapas topográficos

*M^a Ángeles Andreu-Andrés
Miguel García-Casas*

En este trabajo se describe una multitarea de trabajo en grupo en la que, además de compartir responsabilidades en la realización de la misma, los estudiantes de cada grupo autoevalúan la forma de trabajar (evaluación del proceso) y evalúan las aportaciones realizadas por los diferentes grupos durante la lectura de un mapa topográfico, de acuerdo con los criterios acordados por toda la clase en forma de parrillas de evaluación (evaluación del producto). Desde el punto de vista estadístico se indaga sobre la fiabilidad de las variables elegidas para evaluar y comparar, según sea el caso, las puntuaciones otorgadas por el profesor y por los propios alumnos (coevaluación).

Palabras clave: *Evaluación, autoevaluación, coevaluación, trabajo en grupo, trabajo experimental, lectura de mapas topográficos.*

1. Introducción

El tipo de evaluación que se utiliza en el aula influye en el estilo de enseñanza-aprendizaje que se emplea con los alumnos. Así, mientras las preguntas con respuesta múltiple y otros tipos de pruebas responden más a estilos de aprendizaje de reproducción, enfoques como el aprendizaje basado en proyectos, por ejemplo, fomentan la independencia del estudiante unido a diferentes estrategias de aprendizaje (Brown y Pendlebury 1992). El aprendizaje activo aviva la responsabilidad del estudiante sobre su propio progreso además del desarrollo de sus capacidades para aprender a aprender; por lo tanto, darle oportunidades en las que poder ejercer la responsabilidad de su propia enseñanza ayuda al discente a desarrollar

habilidades que le hacen ser un aprendiz más eficaz. Si se quiere cambiar el modo de aprender de un estudiante, también se ha de cambiar la forma de evaluarlo; aprender a aprender es el objetivo más ambicioso e irrenunciable de toda educación, ya que equivale a que el estudiante sea capaz de realizar aprendizajes significativos por sí mismo. Este objetivo recuerda lo importante que es la adquisición de estrategias cognitivas de descubrimiento, exploración, elaboración y organización de la información, sin olvidarnos de la planificación, la regulación y la evaluación de la propia actividad por parte del aprendiz (Dorado, 2001). En consecuencia, la evaluación ha de ir más allá de medir la reproducción de conocimientos (Dochy, Segers y Sluijsmans, 1999).

Los estudiantes, en el ejercicio de su profesión, necesitarán ser capaces de analizar la información que les llega, de resolver problemas conforme surgen y de reflexionar desde un punto de vista crítico sobre su práctica profesional. Las instituciones de educación superior han de contribuir a que los estudiantes mantengan la capacidad de aprender a lo largo de toda su vida. Las metodologías activas, el trabajo en grupo dentro del aprendizaje experiencial y colaborativo, en este caso, constituyen enfoques que favorecen dicho aprendizaje. La investigación en estas metodologías y su evaluación facilitan, sin lugar a dudas, su implantación; si bien no es menos cierto que el éxito en la aplicación de las herramientas didácticas y su evaluación depende del esfuerzo conjunto del alumnado, del personal docente y de la institución en la que se ponen en práctica.

La experiencia de trabajo en grupo de lectura de mapas topográficos (*Mapping and Map Reading*) que aquí se presenta es el resultado de la actividad llevada a cabo en la Escuela Técnica Superior de Ingeniería Geodésica, Cartográfica y Topográfica de la Universidad Politécnica de Valencia, con alumnos de segundo de carrera, durante el primer cuatrimestre del curso 2004-2005, en la asignatura optativa *Inglés Topografía-I*.

Se trata de un grupo de 33 estudiantes divididos en siete equipos, con un nivel de lengua inglesa heterogéneo, desde falsos principiantes a intermedio. Cada equipo se forma de acuerdo con las preferencias de los propios estudiantes, atendiendo a criterios de amistad, afinidad u otros, con un máximo de seis y un mínimo de tres participantes por equipo. Esta actividad se lleva a cabo a mitad de cuatrimestre, cuando los estudiantes ya se conocen entre sí y han realizado las demás actividades del programa de la asignatura en parejas o en pequeños grupos, de modo que cuentan con cierta experiencia para trabajar en equipo.

La finalidad de este estudio es evaluar la experiencia de enseñanza-aprendizaje desde el punto de vista del proceso y del producto, así como el análisis de las variables elegidas para evaluar el producto y comparar las puntuaciones otorgadas por el docente y los propios alumnos, con el fin de averiguar el grado de fiabilidad de la evaluación.

2. Objetivos

De acuerdo con Escobar (2003), las actividades de evaluación repercuten sobre qué y por qué aprenden los alumnos y sobre qué enseña y cómo lo hace el profesor. Desde el punto de vista de la mejora de la enseñanza-aprendizaje, se hace uso de la evaluación formativa para ayudar a los aprendices a que tomen conciencia de los objetivos educativos y puedan ser partícipes, y no meros receptores, en el proceso de instrucción. De este modo llegan a ser responsables de su propio progreso y son capaces de reflexionar sobre el trabajo individual y colectivo, así como de evaluar ambos. Se trata, en definitiva, de ir cambiando en la medida de lo posible la percepción que, en general y siguiendo a MacLellan (2001), tienen los estudiantes de la evaluación en el sentido de que no la utilizan para mejorar su aprendizaje.

Desde el punto de vista de los objetivos interpersonales y académicos, la multitarea de trabajo en grupo que se propone brinda a los estudiantes una serie de actividades sencillas que les ayudan a aprender a trabajar en equipo. En ellas tienen que compartir responsabilidades y resultados de su trabajo con el resto de componentes de su grupo; se trata de tareas que les ayudan a aprender de los demás utilizando la lengua inglesa como instrumento de comunicación y a acostumbrarse a evaluar tanto a los demás como a sí mismos. Para llevar a cabo dichas actividades, los estudiantes precisan hacer uso de las diferentes destrezas comunicativas, además de poner en práctica su habilidad para buscar información relacionada con el tema de trabajo, especialmente en internet.

Respecto a los objetivos de evaluación del *proceso*, se quiere detectar si existen dificultades de acoplamiento entre los componentes de los equipos, además de que los sujetos de cada uno autoevalúen el modo de trabajo de su propio grupo. La autoevaluación se entiende como una herramienta que sirve para resaltar las metas personales del estudiante e identificar el progreso que éste hace a lo largo de unas actividades concretas (Ducasse 2005).

Desde el punto de vista del *producto*¹ el estudio compara cómo puntúan los propios alumnos y el profesor una determinada actividad una vez acordados los criterios de evaluación, de modo que la calificación alcanzada por el representante que hace la defensa oral de un equipo es la misma que obtiene el resto de sujetos de ese equipo de trabajo. De esta forma se fomenta la colaboración entre los componentes del equipo y se comparte la responsabilidad de las tareas. El objetivo de esta práctica evaluadora no se limita a contribuir en la mejora de la habilidad de los estudiantes a la hora de evaluar, sino también a involucrarles, a un nivel más profundo, en las presentaciones orales de sus compañeros. Se trata de que vean la evaluación como una experiencia de aprendizaje, además de como un primer paso para

¹ Presentación oral ante el resto de los equipos de una lectura de mapas de una zona de España.

desarrollar habilidades en el análisis de su propio trabajo de manera realista, con sus puntos fuertes y débiles (MacAlpine 1999).

En los casos de evaluación que se presentan se utiliza una serie de parrillas de evaluación y autoevaluación (cuyas referencias completas aparecen en el apartado bibliográfico) con una serie de variables acordadas previamente con los alumnos, que intentan medir el intercambio oral de información por un lado, la autoevaluación de cada grupo de trabajo por otro y la evaluación de la lectura de un mapa topográfico, por parte de los alumnos y del profesor. Desde el punto de vista estadístico se indaga sobre la fiabilidad de las variables elegidas para evaluar cada una de las tareas y se compara, según el caso, las puntuaciones otorgadas por los propios alumnos y el profesor (coevaluación).

Para evaluar el resultado cada estudiante prepara un breve informe escrito en lengua inglesa en el que se le pide que reflexione sobre las propias tareas y su trabajo en grupo: el procedimiento seguido, los resultados obtenidos, las propuestas de mejora que se le ocurran, etc. Este informe lo valora y corrige únicamente el profesor y forma parte de la calificación final.

3. Modelo de aprendizaje

En términos generales el aprendizaje cooperativo o colaborativo² es un conjunto de actividades en equipo que se desarrollan en el aula tras recibir instrucciones. Quienes conforman los equipos intercambian información, trabajan en una determinada tarea y aprenden a través de la colaboración. De esta manera la pasividad, el individualismo, la competitividad y la memorización se sustituyen por cooperación, responsabilidad para llevar a cabo la tarea asignada, comunicación e intercambio de información entre sujetos. Todo ello unido a otras tareas, siempre presentes en este tipo de aprendizaje, permite el desarrollo de habilidades transversales; la autoevaluación de las acciones y de los resultados obtenidos unido al desarrollo del pensamiento crítico (Johnson y Johnson 1997); la necesaria autonomía del grupo y la responsabilidad de las metas acordadas. El trabajo en grupo es una estrategia de aprendizaje que se ha ido perfilando con los años y en la actualidad es mayoritariamente conocida como trabajo cooperativo.

El constructivismo, si bien no puede decirse que sea un término unívoco, fundamenta su filosofía de aprendizaje sobre la premisa de que el individuo construye su propio conocimiento y entiende el mundo que le rodea a través de la reflexión sobre las experiencias vividas. De esa manera, los constructivistas que defienden las teorías de Vygotsky sobre el aprendizaje consideran que el aprendizaje

² <<http://www.sistema.itesm.mx/va/dide/inf-doc/estrategias/colaborativo.html>>

cooperativo aporta a los estudiantes el andamio que necesitan para mejorar su aprendizaje (Escobar 2002). Según Carretero (1997), se ha comprobado cómo el alumno aprende de forma más significativa cuando lo hace en un contexto de colaboración con sus compañeros, de manera que las discusiones en grupo y el poder de la argumentación entre compañeros que poseen distintos grados de conocimiento sobre un tema estimulan y favorecen el aprendizaje. Ausubel, ya en los años 60, argumenta que la exposición organizada de contenidos por parte del profesor puede ser un instrumento eficaz para que los aprendices consigan una comprensión adecuada, siempre y cuando se tenga en cuenta sus conocimientos previos de la materia, dado que el nuevo conocimiento se asentará sobre el viejo. Una posición mantenida por investigadores constructivistas que pueden considerarse a medio camino entre las aportaciones cognitivas –las de Piaget y las de Vygotsky– sostienen que el intercambio de información entre compañeros con distinto nivel de conocimiento acaba produciendo aprendizaje.

En relación a la evaluación, los constructivistas la defienden como parte del proceso de enseñanza-aprendizaje, de manera que el papel que juega el propio estudiante a la hora de juzgar su propio progreso es fundamental. El papel del docente consistirá en facilitar y dirigir todo el proceso, diseñar y preparar cuidadosamente el marco de actuación de los equipos de trabajo, ya que los alumnos necesitan instrucción para desarrollar las habilidades necesarias para trabajar en equipo, además de tomar decisiones sobre cómo evaluar el rendimiento individual y de grupo.

El modelo de aprendizaje cooperativo que se presenta es el que se ha seguido en la experiencia de evaluación y autoevaluación del trabajo en grupo que se recoge en este trabajo.

4. Materiales y método

La actividad en grupo consta de tres tareas principales en las que cada equipo define, en primer lugar, una serie de conceptos fundamentales relacionados con la cartografía ayudándose de información con vocabulario específico y diferentes direcciones web relacionadas con el tema. A continuación se consensúan aquellos aspectos que, en opinión de todos los equipos, deben tenerse en cuenta a la hora de evaluar los intercambios orales de información de los resultados obtenidos por los diferentes grupos durante esta primera tarea. Los criterios de evaluación se especifican en Tabla 1.

1. Contenido: longitud apropiada, nivel de detalle.	1	2	3	4	5
2. Organización: inicio, introducción.	1	2	3	4	5
3. Lengua y tono de voz: adecuados a la audiencia, gramática correcta, registro adecuado, mínimo uso de “ahs” y “ums”.	1	2	3	4	5
4. Lenguaje corporal: gestos físicos adecuados frente a movimientos innecesarios y nerviosos.	1	2	3	4	5
5. Apoyo visual/transparencias: legible, sencillo, limpio y adecuado para el auditorio, buenos gráficos.	1	2	3	4	5

1 = Escaso

2 - 4 = Medio

5 = Excelente

Tabla 1. Criterios para la evaluación del intercambio oral de información

En la segunda tarea o intercambio oral de información, un representante de cada grupo elegido al azar lleva a cabo, con apoyo del resto de los componentes de su propio grupo, el intercambio de información y la definición de cada uno de los conceptos cartográficos; el resto de equipos también evalúa al representante siguiendo los criterios de evaluación referidos en Tabla 1. Para analizar los resultados obtenidos se calculan los coeficientes de correlación entre las puntuaciones otorgadas a cada grupo en las diferentes variables tenidas en cuenta durante el intercambio oral de información (*contenido, organización, lengua y tono, lenguaje corporal y apoyo visual*); en una primera aproximación se asume que la relación estadísticamente significativa entre estas variables se debería a la relación de cada una de ellas con el fenómeno a valorar: la comunicación en la presentación oral.

Tras estas dos primeras tareas y con objeto de detectar posibles dificultades de acoplamiento entre los componentes de los equipos, se les pide que evalúen el modo de trabajo de su propio equipo hasta el momento; para ello se utiliza una plantilla de autoevaluación del grupo de trabajo³ en la que cada uno de los miembros presentes en ese momento debe evaluar a su propio equipo. La tabla 2 recoge las preguntas traducidas al castellano que sirven para este propósito.

³ Basada en la de Angelo y Cross (1993).

Preguntas	
I	¿Cuántos miembros del grupo han participado más activamente hasta ahora?
II	¿Cuántos miembros del grupo estaban realmente preparados para trabajar en grupo?
III	¿Con qué eficacia ha trabajado tu grupo junto durante estas tareas? (<i>muy buena-buena-adeuada-inadecuada-escasa-ninguna</i>).
IV	Da un ejemplo de algo que has aprendido de tu grupo que probablemente no habrías aprendido por tu cuenta.
V	Da un ejemplo de algo que otros miembros de tu grupo han aprendido de ti que probablemente no lo habrían hecho de otra manera.
VI	Sugiere algún cambio que tu grupo podría hacer y que ayudaría a mejorar el aprendizaje de todos.

Tabla 2. Autoevaluación del grupo de trabajo

Para la tercera y última tarea, cada equipo recibe un mapa topográfico incompleto de una zona; en él tienen que poner en práctica lo aprendido durante las tareas primera y segunda al resolver una serie de cuestiones relacionadas con el mapa incompleto y exponer los resultados obtenidos al resto de equipos. Para la evaluación de esta tarea, por parte de los alumnos y del profesor, se utilizan los criterios recogidos en la tabla 1, para detectar si existen diferencias significativas entre las puntuaciones otorgadas por los discentes y el docente se realiza un análisis de varianza. Las variables tenidas en cuenta durante esta tercera tarea son las mismas que en el intercambio oral de información además del *apoyo visual*, desestimada en la actividad anterior por los estudiantes al considerarla poco relevante para el intercambio de información.

5. Análisis de resultados

5.1. Intercambio oral de información

Durante esta tarea todos los estudiantes presentes puntúan, según los criterios expuestos en la tabla 1, a los compañeros de los demás equipos, nunca a los de su propio grupo. Más que perseguir una puntuación exhaustiva de cada uno de los participantes por parte de los demás grupos, esta actividad se diseña para que los estudiantes se centren en lo que sus compañeros intentan transmitir y cómo lo hacen, además de ir ganando experiencia en la tarea evaluadora. El tratamiento estadístico proporciona una estimación numérica de la eficacia de las variables elegidas; si bien en la tabla de criterios existe un campo dedicado al apoyo visual, cabe mencionar que en esta parte de la actividad se acordó no hacer uso de ningún apoyo visual salvo el que, por necesidades de comprensión, fuera imprescindible unir a la información oral a través de la pizarra.

En el Apéndice se muestra el cómputo de las calificaciones obtenidas por cada uno de los grupos de trabajo en el intercambio oral de información. El número de variables evaluadas entre grupos no es constante; factores como la falta de experiencia en calificar, la valoración de la información oral específica, el esfuerzo de organizar de manera inmediata el discurso oral en una lengua extranjera para añadir o puntualizar la información ofrecida por los demás, la falta de atención, entre otros, pueden ser la causa. La calificación mínima que cada grupo puede obtener por parte de cada evaluador es de cuatro puntos, frente a una máxima de veinte.

Si se parte de la hipótesis de que las variables utilizadas en esta parte del estudio sirven para medir el nivel de intercambio oral de información entre grupos, una manera de confirmar esta hipótesis de trabajo es averiguar si estas variables guardan relación entre sí. Para ello se realiza un análisis de correlación entre las puntuaciones obtenidas por cada grupo en las variables *contenido*, *lengua* y *tono*, *lenguaje corporal* y *organización de la información*, de manera que podamos determinar la existencia o no de relaciones entre estas variables. La tabla 3 recoge los resultados obtenidos en la matriz de correlación, en ella se muestran las correlaciones entre cada par de variables.

En el interior de cada celda aparece el coeficiente de correlación, el tamaño de la muestra entre paréntesis y la significación estadística de la correlación estimada entre cada par de variables. Un valor por debajo de 0,05 indica que hay una relación estadísticamente significativa entre ambas variables comparadas, a un nivel de confianza igual o superior al 95%.

	Contenido	Lengua y tono	Lenguaje corporal	Organización
Contenido		0,7975 (34) 0,0000	0,4488 (34) 0,00078	0,6668 (34) 0,0000
Lengua y tono			0,5448 (34) 0,0009	0,6275 (34) 0,0001
Lenguaje corporal				0,4539 (34) 0,0070

Tabla 3. Matriz de correlación entre variables de intercambio oral de información

Como se aprecia en la tabla 3 las correlaciones más intensas son las referidas a las variables *contenido* con *lengua y tono* y *contenido* con *organización*. Las correlaciones intermedias se dan entre las variables *lengua y tono* con *organización* y *lengua y tono* con *lenguaje corporal*, mientras que las de menor intensidad son *lenguaje corporal* con *organización*, por un lado y *lenguaje corporal* con *contenido*, por otro. Asimismo, los resultados de las correlaciones indican que entre todas ellas hay relaciones estadísticamente significativas.

De igual modo los resultados obtenidos confirman que todas las variables utilizadas para medir el intercambio oral de información están relacionadas entre sí, lo que confirma nuestra hipótesis de trabajo en el sentido de que se puede afirmar que las variables (en este caso, criterios de evaluación) empleadas sirven para medir adecuadamente el intercambio oral de información.

5.2. Autoevaluación del grupo de trabajo

La observación en el aula y una primera autoevaluación de los grupos de trabajo han ayudado a detectar puntos fuertes y débiles de cada grupo. Por lo general, el número de participantes de cada grupo, según los alumnos, coincide con lo observado en el aula por el profesor salvo en uno de los grupos. Tabla 2 recoge las preguntas a las que se refieren los números romanos que aparecen en cada una de las tablas de resultados que se muestran a continuación y que, siguiendo el orden de las mismas, la primera (I) hace alusión al número de integrantes del grupo que han participado más activamente; la segunda (II) es el número de componentes que estaban preparados para trabajar en grupo; la tercera (III) trata de la eficacia con la

que cada grupo ha trabajado; la cuarta (IV) muestra ejemplos de lo que cada uno ha aprendido del grupo; la quinta (V) indica lo que el grupo ha aprendido de uno mismo y la sexta (VI) expone propuestas de cambio para la mejora del aprendizaje de todos los miembros del equipo.

Grupo: 1 Pregunta nº			
<i>I</i>	3	3	3
<i>II</i>	3	3	3
<i>III</i>	Buena	Adecuada	Buena
<i>IV</i>	-	El significado del color morado en los mapas	Mapas con Ortophoto
<i>V</i>	-	-	Mapa topográfico = mapa con ortophoto
<i>VI</i>	-	Mejor organización para realizar las tareas	Miembros del grupo lleguen a tiempo

Tabla 4. Resultados de autoevaluación grupo de trabajo nº 1

Los resultados obtenidos muestran que en el grupo 1 sólo tres de los cuatro componentes trabajan en equipo de manera activa y están dispuestos a trabajar durante las tareas 1 y 2; igualmente coincide el hecho de que son tres de los cuatro miembros los que contestan al cuestionario. El grupo trabaja con una eficacia entre buena y adecuada e incluso se hacen propuestas de mejora para el funcionamiento del grupo en las demás actividades.

Grupo: 2 Pregunta nº		
<i>I</i>	2	2
<i>II</i>	2	2
<i>III</i>	Buena	Buena
<i>IV</i>	Cómo escribir un guión que recoja las principales ideas	Compartir la tarea a realizar
<i>V</i>	Cómo hacer búsquedas en internet	-
<i>VI</i>	-	Todos los miembros del grupo han de ser responsables de las tareas. Respeto hacia todos los miembros

Tabla 5. Resultados de autoevaluación grupo de trabajo nº 2

De los tres miembros que conforman el grupo número 2, dos de ellos trabajan durante las tareas primera y segunda de manera activa y están preparados para el trabajo en grupo. Si bien consideran que la eficacia del trabajo ha sido buena, se observa una clara reivindicación por parte de uno de sus miembros sobre la responsabilidad en la tarea. Asimismo, admiten haber aprendido de los demás a escribir el guión previo al intercambio de información y a ser capaces de compartir su trabajo; de igual manera, uno de sus miembros enseña al resto determinadas técnicas de búsqueda en la red (véase Tabla 5).

Grupo: 3 Pregunta n°				
<i>I</i>	4	4	4	4
<i>II</i>	3	3	4	4
<i>III</i>	Buena	Adecuada	Buena	Buena
<i>IV</i>	Saber organizar información	Saber organizar un grupo de gente	Saber trabajar en grupo	Saberse organizar mejor
<i>V</i>	Saber buscar información	Uso del morado en los mapas	No sé	Saber buscar información
<i>VI</i>	Que todos hablen inglés	Ayudar a otros a hablar en inglés	No tener miedo de hablar en inglés	Comunicarse más en inglés

Tabla 6. Resultados de autoevaluación grupo de trabajo n° 3

En el grupo 3 los resultados obtenidos muestran que los cuatro miembros que forman el grupo participan activamente durante la primera y segunda actividad. No hay acuerdo sobre si todos o sólo tres de ellos están preparados para trabajar en equipo durante la mayor parte del tiempo. Admiten haber aprendido de los demás a organizar la información localizada, a organizarse mejor entre ellos y a trabajar en grupo. Dos de sus miembros han enseñado a los demás a hacer búsquedas en la red y todos son plenamente conscientes de que pueden y deben comunicarse más en inglés a lo largo de todas las actividades (véase Tabla 6).

Grupo: 4 Pregunta n°		
<i>I</i>	2	2
<i>II</i>	2	2
<i>III</i>	Buena	Buena
<i>IV</i>	Gramática	Trabajar en grupo
<i>V</i>	Muchas ideas nuevas	Ortophoto: más precisión
<i>VI</i>	Nada	Hacer deberes

Tabla 7. Resultados de autoevaluación grupo de trabajo n° 4

El grupo número 4 está formado por cinco estudiantes. Según la autoevaluación del grupo (véase Tabla 7) sólo dos trabajan de manera activa y son los únicos que se sienten preparados para trabajar de este modo; la observación de aula corrobora los resultados. Trabajando ambos miembros se sienten bien y consideran que han mejorado su gramática y han aprendido a trabajar en grupo. Creen, asimismo, que han aportado nuevas ideas e información a los demás y consideran necesario que el grupo trabaje fuera de las sesiones de clase.

Grupo: 5 Pregunta n°					
<i>I</i>	5	5	2	5	4
<i>II</i>	4	4	2	5	4
<i>III</i>	Buena	Buena	Buena	Buena	Buena
<i>IV</i>	Mejora de la expresión oral	Vocabulario específico	Algunos conceptos sobre mapas	Saber buscar información	Compartir responsabilidades
<i>V</i>	Un triángulo representa la cima de la montaña	Símbolos de un mapa: triángulo	Información sobre las escalas de un mapa	¿?	El uso de la leyenda de un mapa
<i>VI</i>	Hablar más en inglés	Hablar con todos	Trabajar en equipo	Compartir información	Hablar más en inglés

Tabla 8. Resultados de autoevaluación grupo de trabajo n° 5

De los seis estudiantes que forman el grupo número 5 uno apenas ha contribuido durante las tareas 1 y 2. No obstante, entre los propios miembros del grupo no se ponen de acuerdo sobre el número de estudiantes que lo forman: entre cinco y dos, hecho que puede deberse a que no todos tienen la sensación de colaborar en

el grupo. De igual modo existe una discrepancia importante a la hora de considerar cuántos miembros del grupo están preparados para el trabajo en grupo, a pesar de ello creen que la efectividad con la que el grupo ha trabajado ha sido buena. Consideran que gracias a los demás han mejorado su expresión oral, han aprendido el significado de determinados símbolos que aparecen en los mapas, cómo buscar información y a compartir responsabilidades. Salvo uno de los cinco miembros, que no entendió la pregunta IV, los demás consideran que han aportado algo a los demás. Asimismo, son conscientes de que deben hablar más en inglés entre ellos (véase Tabla 8).

Grupo: 6 Pregunta nº		
<i>I</i>	1	1
<i>II</i>	2	2
<i>III</i>	Inadecuada	Inadecuada
<i>IV</i>	Trabajar en grupo es más sencillo que individualmente.	He aprendido a preparar un pequeño informe oral
<i>V</i>	Terminología específica	Nada aunque he escrito mucho
<i>VI</i>	Mayor participación	La responsabilidad no está compartida

Tabla 9. Resultados de autoevaluación grupo de trabajo nº 6

Los resultados obtenidos en el grupo 6 muestran que sólo uno de los seis miembros participa activamente en las tareas 1 y 2. La observación en el aula confirma este hecho si bien los dos estudiantes que contestan al cuestionario consideran que ambos están preparados para trabajar en equipo. Uno de ellos señala que trabajar en grupo es más sencillo que hacerlo solo, si bien los dos estudiantes que contestan el cuestionario reconocen que la eficacia del grupo ha sido inadecuada; consideran que han aprendido terminología y puesto en práctica su expresión escrita. Es evidente que las responsabilidades no están compartidas en este grupo porque la mayor parte del tiempo ha estado trabajando de firme sólo una persona. Estos datos unidos a las observaciones de aula pueden ayudar al docente a decidir cambios en el grupo (véase Tabla 9).

Grupo: 7 Pregunta n°					
<i>I</i>	6	4	5	2	2
<i>II</i>	2	0	1	2	2
<i>III</i>	Buena	Adecuada	Buena	Buena	Adecuada
<i>IV</i>	-	Preparar información para exponer con sencillez	-	Diferentes tipos de mapas	Diferencias entre mapa y ortophoto
<i>V</i>	-	Búsqueda de información	Terminología específica	Cosas necesarias para ortophoto	-
<i>VI</i>	Más tiempo	Más tiempo	Mejorar la comunicación	Miembros deben ir a clase	Mayor participación

Tabla 10. Resultados de autoevaluación grupo de trabajo n° 7

El grupo número 7 está formado por seis estudiantes aunque por las opiniones de los que contestan a la encuesta no hay acuerdo sobre el número de sus miembros: entre dos y seis. De todos ellos parece que sólo dos están preparados para trabajar en grupo. Sin embargo, consideran que la eficacia del grupo ha sido entre buena y adecuada. Tres de ellos afirman haber aprendido de los demás y, a su vez, haber enseñado algo. Si bien dos de ellos consideran que las tareas necesitan una asignación de tiempo mayor, otros dos reclaman una mayor participación por parte de los demás miembros del grupo; las observaciones de aula coinciden con estas dos opiniones (véase Tabla 10).

A la vista de los resultados y de acuerdo con Orsmond, Merry y Reiling (1997) se considera que la autoevaluación es una herramienta educativa muy útil que ayuda a los estudiantes a alcanzar sus objetivos de aprendizaje y a ejercer una práctica reflexiva de la misma. Como se trata de una actividad poco extendida en la educación universitaria no es de extrañar que los estudiantes estén poco acostumbrados a reflexionar sobre el trabajo realizado, hecho que se constata en los resultados, especialmente cuando admiten haber aprendido algo de sus compañeros. La sorpresa de haber aprendido algo de los demás al trabajar en equipo se percibe en la observación de aula y en los casos en que afirman no saber qué pueden haber aprendido de los demás, una muestra más de lo poco habituados que están a reflexionar sobre las tareas llevadas a cabo.

Además de encontrar una buena dosis de sinceridad en las respuestas, se aprecia cierto espíritu crítico, gracias al cual se perciben los puntos débiles de cada grupo: falta de responsabilidad de algunos miembros, la participación de otros, el reconocimiento de que en determinados equipos se habla español a la mínima oportunidad. Teniendo en cuenta las personas que componen cada equipo, se puede constatar que aquellos estudiantes con un trabajo mejor realizado son más autocríticos consigo mismos que con los demás; al contrario de lo que sucede con aquellos con unos resultados más pobres, conclusión que Orsmond y otros (1997) ya mostraban en su trabajo.

5.3. Lectura de mapas

En esta tarea cada grupo utiliza los instrumentos que considera más adecuados para calcular y dar respuesta a cada una de las cuestiones relacionadas con el mapa incompleto y exponer los resultados obtenidos al resto de grupos. En la evaluación de lectura de mapas intervienen las variables *contenido*, *organización*, *lengua y tono*, *lenguaje corporal* y *apoyo visual*. El resultado de esta tarea es evaluado por los componentes de los grupos restantes y por el profesor a fin de comparar, con posterioridad, los resultados de ambas.

	Calificación de los alumnos (Muestra 1)	Calificación del profesor (Muestra 2)
Valores obtenidos	217	29
Media de puntuaciones	3,182	3,0
Varianza	0,801203	0,803571
Desviación estándar	0,8951	0,896421
Puntuación mínima	1,0	2,0
Puntuación máxima	5,0	5,0

Tabla 11. Resumen estadístico muestras 1 y 2

La muestra 1 de la tabla 11, correspondiente a las notas otorgadas por los alumnos, cuenta con 217 valores que oscilan en un rango de uno a cinco en cada una de las cinco variables mencionadas en el párrafo anterior. Respecto a la muestra 2, correspondiente a las notas dadas por el profesor, se cuenta con 29 valores que oscilan en un rango de entre dos y cinco. Esta tabla ofrece el resumen estadístico de las calificaciones de ambas muestras respecto a las medias de calificación, varianza, desviación típica, puntuaciones mínimas y máximas de los estudiantes y el

profesor. De acuerdo con Magin (2001) cuando se comparan las puntuaciones otorgadas sobre el proceso del trabajo en grupo por los alumnos y un único profesor, se lleva a cabo un análisis de varianza; si bien en este caso las puntuaciones comparadas para saber su fiabilidad no se refieren al proceso sino al producto del trabajo en grupo.

La media de las puntuaciones entre estudiantes y profesor es prácticamente idéntica con una diferencia de 0,1 puntos a favor de los estudiantes. Sin embargo, las puntuaciones de los estudiantes oscilan entre 1 y 5 puntos, cuando el profesor lo hace entre 2 y 5 puntos; de ello se deduce que los alumnos son algo más severos que el profesor. Estudios como los de Oldfield y MacAlpine (1995), Fry (1990) y Rushton (1993) señalan que, al contrario de lo que cabría esperar, las puntuaciones otorgadas por los discentes son significativamente iguales a las otorgadas por los docentes, por lo que la evaluación hecha por ambos es igualmente válida.

El análisis de varianza muestra un valor para F de 1,09 para 1 y 244⁴ grados de libertad, lo que indica que no hay diferencias estadísticamente significativas a un nivel igual o superior al 95% entre las medias de las calificaciones otorgadas por los alumnos y por el profesor en la lectura de mapas topográficos. No obstante, siendo consecuentes con el análisis, se puede asumir que el conjunto de calificaciones otorgadas por los alumnos puede representar al conjunto universal de “calificaciones de alumnos” mientras que el conjunto de calificaciones del profesor es una muestra del “conjunto universal de calificaciones que pudiera dar ese único profesor” y no “el conjunto universal de profesores”, ya que no se puede asumir como punto de referencia las puntuaciones otorgadas por un solo profesor (Magin y Helmore 2001). Dicho de otro modo y siguiendo a Magin (2001), el escenario ideal para llegar a estimar la fiabilidad de las calificaciones del profesor debe contar con dos o más profesores que puntúen de acuerdo con unos criterios consensuados; si ello no es posible, el análisis de varianza para comparar las calificaciones de los alumnos y las de un único profesor, como este estudio recoge, puede considerarse una técnica novedosa y útil pero de la que no hay que ignorar sus limitaciones.

Los resultados obtenidos confirman, no obstante, la aseveración de autores como Dochy, Segers y Sluijsmans (1999) en el sentido de que involucrar a los propios estudiantes en el proceso evaluativo cuenta con un grado de validez y fiabilidad significativa.

⁴ 1 grado de libertad (2 grupos menos 1); 244 grados de libertad (246 valores menos n° grupos analizados).

6. Conclusiones

El estudio que aquí se presenta supone un primer paso dentro del complejo campo de la evaluación del trabajo en grupo. No sólo se aportan nuevos conocimientos al docente sino también a los discentes al plantearse (por primera vez en el caso de los alumnos con los que se realiza el experimento) qué han aprendido de los demás compañeros y qué han aprendido los demás de ellos a lo largo de una de las tareas. Han sido capaces de detectar los puntos fuertes y débiles de sus respectivos grupos y ofrecer propuestas de mejora. Se han consensuado variables y criterios con las que valorar el intercambio oral de información, cuyas relaciones han resultado ser estadísticamente significativas. Han evaluado la lectura de mapas topográficos realizada por sus compañeros y han otorgado unas calificaciones en las que no hay diferencias estadísticamente significativas con las dadas a los mismos alumnos por el propio profesor, tras utilizar una plantilla de evaluación común.

Queda un largo camino por recorrer que en este caso debería seguir por plantearse ampliar la muestra de la primera y segunda tarea y contrastar resultados. Aumentar la muestra de profesores que evalúen la lectura de mapas de los alumnos sería otra acción a perseguir, de manera que los resultados que se obtengan al comparar una muestra y otra puedan ser generalizables.

Bibliografía

- Angelo, T.K. y Cross, K.P. (1993, 2^a ed.). *Classroom assessment techniques*. San Francisco: Jossey-Bass.
- Ausubel, D.P., Novack, J.D. y Hanesian, H. (1983). *Psicología educativa*. México: Trillas.
- Brown, G. y Pendlebury, M. (1992). *Assessing active learning, Part 1. Core materials*. Sheffield: CVCP Universities.
- Carretero, M. (1997). ¿Qué es el Constructivismo? *Desarrollo cognitivo y aprendizaje. Constructivismo y educación*. México: Progreso.
- Dochy, F., Segers, M. y Sluijsmans, D. (1999) The Use of self-, peer and co-assessment in higher education. *Studies in Higher Education*, 24(3): 331-350.
- Dorado, C. (2001). *Aprender a aprender. Estrategias y técnicas*. “La Fuente Psicológica”. <http://www.xtec.es/-cdorado/cdora1/esp/psicolog.htm> [ref. 12 noviembre 2001].

- Ducasse, A.M. (2005). La autoevaluación como parte de la nota semestral. *Revista RedELE*, nº 0. <http://www.sgci.mec.es/redele/revista/ducasse.htm> [ref. 28 abril 2005].
- Escobar, R.M^a (2002). *Trabajo en grupo y aprendizaje cooperativo*. http://redescolar.ilce.edu.mx/redescolar/biblioteca/articulos/pdf/trabajo_en_grupo.pdf [ref. 24 abril 2004].
- Escobar Urmeneta, C. (2003). La evaluación de la interacción oral. *Mosaico*, Consejería de educación en Bélgica, Países Bajos y Luxemburgo, 10: 18-25.
- Fry, S.A. (1990). Implementation and evaluation of peer marking in higher education. *Assessment & Evaluation in Higher Education*, 15: 177-189.
- Johnson, D.W. y Johnson, F.P. (1997). *Joining together: Group theory and group skills*. Needham Heights, MA: Allyn & Bacon.
- MacAlpine, J.M.R. (1999). Improving and encouraging peer assessment of student presentations. *Assessment & Evaluation in Higher Education*, 24(1): 15-25.
- MacLellan, E. (2001). Assessment for learning: The differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education*, 26(4): 307-318.
- Magin, D.J. (2001). A novel technique for comparing the reliability of multiple peer assessments with that of single teacher assessments of group process work. *Assessment & Evaluation in Higher Education*, 26(2): 139-152.
- Magin, D. J. y Helmore, Ph. (2001). Peer and teacher assessments of oral presentation skills: How reliable are they? *Studies in Higher Education*, 26(4): 288-298.
- Oldfield, K.A. y Macalpine, J.M.K. (1995). Peer and self-assessment at the tertiary level – An experiential report. *Assessment & Evaluation in Higher Education*, 20: 125-132.
- Orsmond, P., Merry, S., Reiling, K. (1997). A study in self-assessment: Tutor and students' perceptions of performance criteria. *Assessment & Evaluation in Higher Education*, 22(4): 357-369.
- Rushton, C., Ramsey, P. y Rada, R. (1993). Peer assessment in a collaborative hypermedia environment: A case study. *Journal of Computer-based Instruction*, 20: 75-80.

Apéndice. Resultados del intercambio oral de información tras la tarea primera

Puntuaciones obtenidas por el grupo 1.

Intercambio información	G1	G1	G1	G1	G1	G1
<i>1. Contenido</i>	4	5	3	4	4	3
<i>2. Organización</i>	3	4	2	5	5	4
<i>3. Lengua y tono de voz</i>	3	5	3	3	4	4
<i>4. Lenguaje corporal</i>	2	4	4	4	4	4
TOTAL	12	18	12	16	17	15

Media: 15 puntos

Puntuaciones obtenidas por el grupo 2.

Intercambio información	G2	G2	G2	G2
<i>1. Contenido</i>	5	3	3	4
<i>2. Organización</i>	4	3	4	3
<i>3. Lengua y tono de voz</i>	5	3	3	3
<i>4. Lenguaje corporal</i>	4	4	4	3
TOTAL	18	13	14	13

Media: 14,5 puntos

Puntuaciones obtenidas por el grupo 3.

Intercambio información	G3	G3	G3	G3	G3
<i>1. Contenido</i>	3	5	4	4	3
<i>2. Organización</i>	4	4	3	4	4
<i>3. Lengua y tono de voz</i>	4	5	3	4	3
<i>4. Lenguaje corporal</i>	2	4	2	5	4
TOTAL	13	18	12	17	14

Media: 14,8 puntos

Puntuaciones obtenidas por el grupo 4.

Intercambio información	G4	G4	G4	G4	G4
<i>1. Contenido</i>	1	4	3	3	4
<i>2. Organización</i>	2	3	4	2	3
<i>3. Lengua y tono de voz</i>	1	4	3	3	4
<i>4. Lenguaje corporal</i>	1	3	3	2	3
TOTAL	5	14	13	10	14

Media: 11,2 puntos

Puntuaciones obtenidas por el grupo 5.

Intercambio información	G5	G5	G5	G5	G5
<i>1. Contenido</i>	4	2	3	4	2
<i>2. Organización</i>	4	2	3	3	3
<i>3. Lengua y tono de voz</i>	4	3	3	3	3
<i>4. Lenguaje corporal</i>	3	3	2	2	3
TOTAL	15	10	11	11	11

Media: 11,6 puntos

Puntuaciones obtenidas por el grupo 6.

Intercambio información	G6	G6	G6	G6	G6
<i>1. Contenido</i>	1	4	1	2	2
<i>2. Organización</i>	1	3	1	2	3
<i>3. Lengua y tono de voz</i>	1	4	2	3	3
<i>4. Lenguaje corporal</i>	1	3	3	3	2
TOTAL	4	14	7	10	10

Media: 9 puntos

Puntuaciones obtenidas por el grupo 7.

Intercambio información	G7	G7	G7	G7
<i>1. Contenido</i>	3	3	3	3
<i>2. Organización</i>	3	2	2	3
<i>3. Lengua y tono de voz</i>	2	3	3	4
<i>4. Lenguaje corporal</i>	2	4	4	2
TOTAL	10	12	12	12

Media: 11,5 puntos



Experiencias de evaluación de trabajo en grupo en el área de matemáticas

María José Pérez Peñalver

Este trabajo describe una primera aproximación a la evaluación del trabajo en grupo en dos asignaturas de matemáticas de los primeros cursos de Ingeniería Civil. Si bien las actividades grupales que se plantean son, en el fondo y en la forma, totalmente diferentes, comparten el objetivo común de introducir a los estudiantes en la estrategia del trabajo colaborativo. Se han utilizado diferentes técnicas para evaluar tanto el producto como el proceso del trabajo realizado por los grupos, actividad en la que se ha involucrado a los alumnos. Los resultados obtenidos permiten extraer conclusiones respecto a cómo mejorar la organización de los trabajos en grupo y a cómo se sienten los alumnos al realizar este tipo de tarea.

Palabras clave: *Aprendizaje colaborativo, evaluación de grupos, evaluación del proceso, evaluación entre iguales, matemáticas en carreras técnicas.*

1. Introducción

Durante los últimos cursos académicos se ha introducido diversas experiencias de trabajo en grupo en dos asignaturas de matemáticas de los primeros cursos de Ingeniería Civil. La primera asignatura es Cálculo de primer curso, troncal, anual y con clases numerosas (alrededor de 90-100 alumnos), la segunda, Matemáticas Asistidas por Ordenador, optativa, cuatrimestral de segundo curso, donde el tamaño medio del grupo es pequeño (de 12 a 20 alumnos aproximadamente).

El planteamiento en estas asignaturas es totalmente diferente, puesto que el diseño de la asignatura de segundo curso desde el inicio estuvo pensado para unas clases más dinámicas, menos rígidas y con un tercio de la asignatura dedicado a un

proyecto que se trabaja en grupo¹; la asignatura de primer curso se desarrolla mayoritariamente mediante clase magistral o clase magistral participativa.

Si bien las actividades grupales que se plantean en ambas asignaturas son, en el fondo y en la forma, totalmente diferentes, comparten el objetivo común de introducir a los estudiantes en la dinámica del trabajo cooperativo. Sobre la conveniencia de introducir actividades de este tipo en las clases de matemáticas se pronuncian diversos autores como Artzt y Newman (1990), MacBean, Graham y Sangwin (2003), Challis et al. (2003). Sus argumentos se basan, por un lado, en la necesidad de la sociedad y del mercado de trabajo de contar con profesionales con ciertas habilidades como las comunicativas, transferir la experiencia de un área a otra, independencia, saber trabajar en equipo y, por otro lado, en la conveniencia de utilizar trabajos en grupo para que se produzca un aprendizaje de nivel cognitivo superior.

2. Trabajo en grupo

El tamaño de los grupos de trabajo se han establecido en ambas asignaturas a elección libre de los estudiantes, siendo de cinco componentes en la asignatura troncal de Cálculo y de tres en la optativa, Matemáticas Asistidas por Ordenador. Los objetivos difieren según la asignatura.

Para la asignatura de Cálculo se proponen dos trabajos, uno para cada cuatrimestre, que consisten en la resolución y redacción de cinco problemas tipo de un tema del programa. Estos dos temas no se tratan de manera habitual en clase con una clase magistral, sino que los alumnos, con directrices del profesor, los preparan junto con el resto compañeros de grupo. De esta forma se pretende que los estudiantes manejen bibliografía de la asignatura con diferentes enfoques, bien la facilitada por el profesor o la que recopilan ellos mismos. Por tanto, se persigue favorecer el autoaprendizaje y la búsqueda de información. Otro objetivo que se plantea es que los alumnos se enfrenten a la redacción correcta de contenidos de matemáticos, es decir, que sean capaces de expresarse y hacerse entender, particularmente al utilizar el lenguaje matemático. Se pretende también que todos los integrantes del equipo puedan defender oralmente ante el profesor las soluciones propuestas a los ejercicios que se les plantea. En cuanto a los objetivos de la evaluación, se intenta que los alumnos puedan aprender de sus errores, por lo que, durante el desarrollo de la tarea, en diferentes momentos del proceso, el profesor

¹ Para más información sobre la asignatura Matemáticas Asistidas por Ordenador consultar García Raffi L. M., Pérez Peñalver M.J., Sánchez Pérez E. A., Sánchez Pérez J. V. (2000)).

recoge las soluciones obtenidas hasta el momento y devuelve a cada equipo indicaciones y sugerencias respecto a lo que han hecho, guiando así el trabajo.

Para la segunda asignatura, Matemáticas Asistidas por Ordenador, el trabajo en grupo tiene lugar en el último tercio de la asignatura (1.5 créditos), donde los objetivos específicos se inician con la capacidad de enfrentarse a un problema-proyecto abierto multidisciplinar, de una longitud considerable. Al unir los conocimientos y las capacidades de tres personas, los estudiantes proponen una solución (o varias), que deben defender oralmente con ayuda de los medios que ellos mismos decidan. En este caso, se encuentran con tareas y obstáculos muy diferentes: búsqueda de información en diferentes contextos, resolución de problemas matemáticos más o menos complejos, traslación de los resultados a otras disciplinas, debates y reparto de tareas en el grupo, elección de cómo llevar a cabo una defensa efectiva de la solución planteada. De esta forma, los estudiantes comparten, profundizan y aplican un mayor número de contenidos matemáticos. Este tipo de tareas planteadas introducen a los alumnos en el Aprendizaje Basado en Proyectos o PBL (*Problem Based Learning*), metodología expuesta en Robert, Stewart, Quinney y Hibberd (2002), en López-Real y Rita Chan (1999), en la DIDE del Instituto Tecnológico y de Estudios Superiores de Monterrey o en la denominada modelización matemática propuesta por Gómez Urgellés (2003).

En las dos asignaturas se tienen objetivos comunes, objetivos específicos del trabajo colaborativo: desarrollar la comunicación, la cooperación, el intercambio de ideas, la negociación, la resolución de conflictos, la toma de decisiones y, con todo, llevar a cabo un aprendizaje de nivel cognitivo más alto, más significativo y más solidario.

También se pretende que los estudiantes se involucren en el proceso de evaluación, que reflexionen sobre el trabajo realizado, sobre cómo han ejecutado la tarea y sobre su forma de trabajar juntos, de modo que sean capaces de evaluar el proceso de la tarea grupal, para lo que se llevan a cabo ejercicios de evaluación entre iguales y de autoevaluación en cada uno de los trabajos planteados.

3. Evaluación del producto y del proceso

En la asignatura de Cálculo el producto que se obtiene en las diferentes etapas del trabajo en grupo se evalúan por el profesor y constituyen el 10% de la calificación total. El alumno interviene en la evaluación del proceso. Cada trabajo atraviesa diferentes momentos de evaluación:

- Tres sesiones en el aula, separadas entre ellas por una semana, para que los alumnos se vayan iniciando en la resolución de los problemas. En cada sesión el profesor recoge los resultados del trabajo realizado y lo devuelve al equipo la misma semana con las sugerencias pertinentes.

- En la cuarta semana se realiza una tutoría obligatoria con cada grupo, con una duración de aproximadamente 30 minutos. En ella se consultan dudas referentes al trabajo y los alumnos evalúan la forma de trabajar en grupo mediante una parrilla de criterios consensuados en el aula (ver Apéndice A). Por observación, el profesor también anota las peculiaridades que considera convenientes sobre cada equipo.
- En la quinta semana los equipos entregan el trabajo por escrito; durante esa misma semana el profesor devuelve la corrección del trabajo y las sugerencias para que rectifiquen los ejercicios incompletos o mal resueltos.
- En la sexta semana cada equipo entrega el trabajo definitivo.

Cada componente del grupo completa las diferentes parrillas (consultar Apéndice A) en dos momentos de evaluación. El primer momento se realiza mediante la parrilla 2 (autoevaluación). En la primera tabla de la segunda parrilla, cada alumno reparte 100 puntos en cada fila entre los miembros de su grupo incluido el mismo si el grupo lo constituyen 5 personas, 120 puntos si son 6 personas; no se admite una puntuación por igual de los integrantes. En la segunda tabla de la segunda parrilla, cada alumno puntúa de uno a cinco a sí mismo y a los demás. El segundo momento de evaluación se lleva a cabo con la parrilla 4 (evaluación final del proceso) donde el alumno evaluador elige entre los ítems establecidos y comenta la experiencia de trabajar en grupo.

Para el trabajo en grupo en la asignatura de Matemáticas Asistidas por Ordenador, el cual consiste en un problema-proyecto abierto y cuya valoración proporciona el 50 % de la calificación total de la asignatura, se tienen los siguientes momentos de evaluación:

- Con una tormenta de ideas en el aula se generan criterios para evaluar el producto y el proceso del trabajo en grupo. Después, cada alumno elige cuatro criterios para el producto y cuatro para el proceso (voto múltiple individual). Posteriormente, todos los alumnos se reúnen en tres grupos, reflexionan sobre lo que han escrito y eligen los cuatro o cinco criterios que consideran más idóneos para valorar el producto y otros tantos para evaluar el proceso. Por último, se reúnen portavoces de cada grupo, ponen en común los resultados y deciden los criterios definitivos para el producto y para el proceso; estos criterios se hacen públicos a toda la clase. Los criterios para evaluar el proceso se obtienen de esta forma, sin embargo, el profesor interviene en los criterios para evaluar el producto.

- Llegado el momento de concluir el proyecto y su defensa ante un grupo de profesores (generalmente dos) y ante los compañeros, se dispone de dos parrillas de evaluación (consultar Apéndice B), una para el producto y otra para el proceso. El producto lo evalúan los profesores mientras que el proceso lo evalúa el propio grupo.
- Hay una última entrevista, de carácter optativo, de cada grupo con el profesor, en la que se comentan los resultados obtenidos y se realiza una reflexión del trabajo y de la asignatura.

Cada miembro del grupo rellena la parrilla referente al proceso (consultar Apéndice B). En relación al producto, se evalúa por dos o más profesores. En la primera tabla, cada alumno debe repartir 100 puntos en cada fila entre los miembros de su grupo (incluido él mismo) y está establecido el no puntuar a todos los miembros por igual. En la segunda tabla cada profesor puntúa con 2, 4, 6, 8 o 10 en cada apartado como se indica en la parrilla.

4. Análisis de resultados

4.1. Cálculo

En esta asignatura se ha contado con 16 grupos, casi todos formados por cinco personas. Los resultados obtenidos se resumen en Tabla 1.

La evaluación compartida: investigación multidisciplinar

	PRODUCTO 1	PRODUCTO 2	PRODUCTO FINAL	PROCESO
Grupo 1	1	0,9	Bueno	Bueno. Un componente menos implicado. El componente menos implicado abandona.
Grupo 2	1	0,6	Mediocre	Mediocre. Un componente no colabora y es eliminado del grupo.
Grupo 3	1	0,9	Bueno	Bueno.
Grupo 4	0,95	1	Bueno	Bueno.
Grupo 5	0,8	0,75	Mediocre	Malo. Dos componentes del grupo son eliminados. Muy mal reparto de tareas.
Grupo 6	0,8	0,4	Mediocre	Mediocre. Un componente abandona el grupo.
Grupo 7	0,75	0,8	Mediocre	Mediocre.
Grupo 8	0,65	0,8	Mediocre	Bueno. Un componente abandona el grupo.
Grupo 9	0,45	0,9	Mediocre	Bueno.
Grupo 10	0,9	0,9	Bueno	Bueno.
Grupo 11	0,9	0,6	Mediocre	Bueno. Existe un abandono y un traslado por problemas con el grupo anterior.
Grupo 12	1	1	Bueno	Bueno.
Grupo 13	1	0,9	Bueno	Bueno.
Grupo 14	0,85	0,7	Mediocre	Bueno.
Grupo 15	1	1	Bueno	Bueno.
Grupo 16	1	0,95	Bueno	Mediocre. Un componente no integrado. Cambio de grupo del componente no integrado.

Tabla 1. Resumen de los resultados obtenidos en la asignatura de Cálculo

Si resumimos los resultados obtenidos, hay 7 grupos trabajan bien el proceso y el producto (44%); 4 grupos que trabajan bien el proceso y mediocre el producto (25%); un grupo que trabaja bien el producto y mediocre el proceso (6%); 3 grupos obtienen mediocre el proceso y el producto (19%) y un grupo lleva mal el proceso y alcanza un producto mediocre (6%).

Al empezar el segundo cuatrimestre, y por tanto después del primer trabajo en grupo, se les formula a los alumnos una serie de preguntas, entre ellas cómo se sienten con el trabajo en grupo y si lo consideran eficaz para aprender (ver Figura 1).

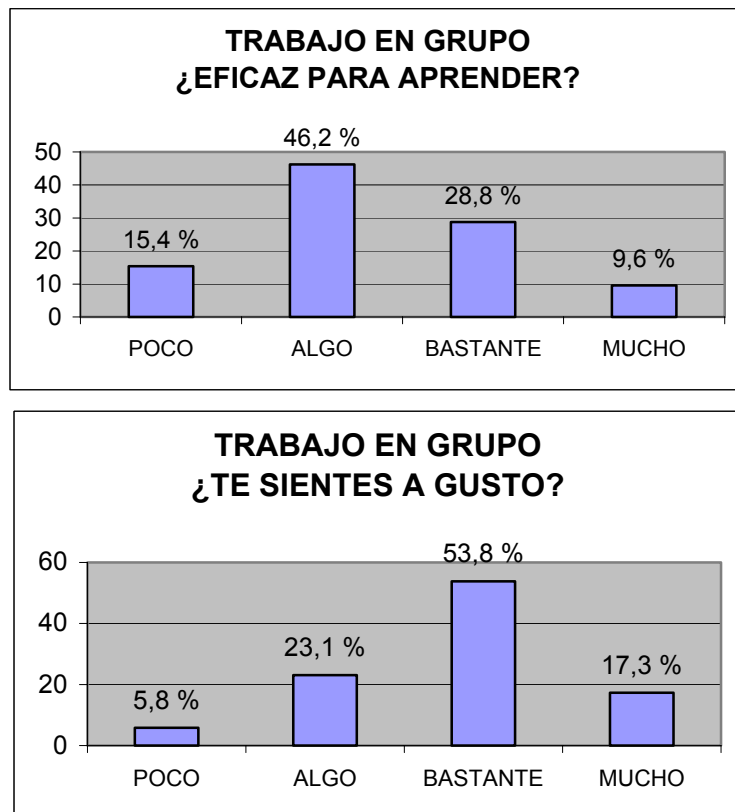


Figura 1. Resultados de la encuesta de opinión en la asignatura de Cálculo

Los resultados de la encuesta muestran que los estudiantes se sienten bastante a gusto con el trabajo en grupo, sin embargo la opinión media es que lo consideran moderadamente eficaz para aprender.

Efectividad	Muy productivo y efectivo 23,95%	Moderadamente efectivo 76,05%	Poco efectivo 0%	
Experiencia de trabajo en grupo	Muy buena 45,71%	Satisfactoria 45,71%	Decepcionante 8,58%	
Calidad personal de las contribuciones al grupo	Excelente 2,81%	Muy buena 32,39%	Buena pero con posibilidad de mejora 56,33%	Pobre 8,75%

Tabla 2. Resultados de la encuesta de todos los grupos

La tabla 2 contiene los resultados globales de la valoración de todos los grupos, que amplían y apoyan la información que aportan los resultados de la figura 1.

4.2. Matemáticas Asistidas por Ordenador

A pesar de que se formaron más grupos, sólo se tienen datos de siete, ya que la última reunión con los grupos fue optativa. Los resultados completos se resumen en la tabla 2.

	PRODUCTO	PROCESO
Grupo 1	6,5	Malo. Cuatro componentes.
Grupo 2	8,5	Bueno.
Grupo 3	8,2	Bueno. Dos componentes.
Grupo 4	7,8	Bueno.
Grupo 5	7,4	Bueno.
Grupo 6	6	Bueno.
Grupo 7	7,2	Bueno.

Tabla 2. Resumen de los resultados obtenidos en Matemáticas Asistidas por Ordenador

Si analizamos los resultados obtenidos tenemos 5 grupos que trabajan bien el proceso y el producto (72%); un grupo que lleva bien el proceso pero mediocre el producto (14%) y un grupo que ejecuta mal el proceso y obtiene un producto mediocre (14%).

5. Conclusiones

Cálculo

Respecto a los objetivos trazados, se quiere señalar que es un tanto decepcionante que no todos los grupos hayan realizado bien los cinco problemas, aun cuando incluso ellos reconocen tener suficientes facilidades para llevarlos a buen término. Respecto a cómo perciben la experiencia de trabajar en grupo, exceptuando a los grupos que no han funcionado bien y las personas individualistas reacias en cualquier circunstancia a trabajar con otras personas, se observa que les resulta una buena experiencia que volverían a repetir. Piensan, en general, que podrían haber contribuido más al grupo y que el trabajo podría haber sido más efectivo. Esta idea de no efectividad está muy extendida en la sociedad, se tiene la sensación de “perder el tiempo” en las reuniones. Con esta experiencia se constata que los alumnos recién llegados a la universidad no saben llevar a cabo reuniones eficientes y satisfactorias. A este respecto, cabe señalar que muchas investigaciones sobre aprendizaje colaborativo y cooperativo han propuesto que los beneficios de realizar actividades en grupo no se ponen de manifiesto de inmediato si no que se obtienen normalmente a más largo plazo.

Por ejemplo, usar en los primeros cursos técnicas de trabajo en grupo puede ayudar a los estudiantes a reflexionar más sobre su trabajo, lo cual puede ser aprovechado más adelante en cursos superiores (MacBean, Graham, y Sangwin, 2003). Estos autores proponen que a los estudiantes se les debe facilitar cierta información cuando empiezan a trabajar en grupo:

- Las tareas y detalles de la forma de presentación por escrito.
- La composición del grupo.
- Detalles y ayudas que podrían apoyar el trabajo del grupo.
- Consejos sobre cómo trabajar en grupo.
- Criterios de evaluación de la tarea.

Como resultado de la experiencia de este curso, debemos tomar decisiones para cursos posteriores si queremos que mejoren los resultados. Se estiman oportunas las siguientes propuestas:

1. Compatibilidad de horarios: en la formación de los grupos la única condición que se pedirá es que puedan reunirse todos los miembros una vez a la semana durante el primer cuatrimestre y durante el segundo.

2. Información previa al trabajo en grupo: plantear los problemas que puedan surgir durante el proceso.
3. Guiar más el trabajo: establecer calendarios de entregas parciales para conseguir un trabajo continuo y más organizado.

Matemáticas Asistidas por Ordenador

En esta asignatura, que lleva impartiendo aproximadamente siete cursos, la tendencia general es que los grupos de trabajo funcionen bastante bien durante todo el proceso. Incluso, grupos en los que sus componentes no tienen buen nivel en matemáticas han llegado a un nivel de profundización sobre los contenidos de su proyecto por estar muy motivados. Las razones fundamentales del éxito son el número de personas por grupo (3) que hace más dinámico el proceso, el número de alumnos por clase (12-20), el hecho de que la mayoría del trabajo se realice en horas de clase con apoyo del profesor y que la tarea a realizar les resulta más motivadora de lo habitual.

Durante las reuniones optativas con los grupos para reflexionar sobre los resultados (producto) y sobre el trabajo realizado (proceso), los alumnos hacen comentarios que respaldan estas conclusiones:

- “Lo que más me ha gustado de hacer el trabajo es que por primer vez he visto una aplicación real de las matemáticas”, “... me ha dado la sensación de que estaba haciendo un proyecto de verdad”.
- “Al utilizar la teoría aprendida en clase durante el trabajo, creo que se nos han quedado mucho mejor los conceptos”.
- “... me es bastante incómodo hacer presentaciones en público, pero creo que haciendo este tipo de trabajos y pequeñas presentaciones se ayuda bastante a coger poco a poco un poco más de confianza y mejorar la forma de hablar”, “...la experiencia ha sido muy positiva, pues hemos aprendido a enfrentarnos a una presentación y exposición del trabajo, cosa que es algo nuevo para nosotros”, “... prueba para superar el miedo escénico... cosa a la que hay que acostumbrarse”.
- “...a parte de conocer a nueva gente, hemos intercambiado conocimientos y todos hemos aprendido algo del resto, además es un inicio para aprender a trabajar en grupo, que es a lo que nos dedicaremos en la actividad profesional”. “Trabajando en grupo se aprenden más, una cosa importante es que lo que a uno no se le ocurre se le ocurre al otro”.
- “... como pudimos hacer buena parte del trabajo en horas de clase, no nos quitó demasiado tiempo”.

- “Respecto a la forma de evaluación empleada creo que ha sido buena”, “... encontrarse con una asignatura de este tipo es algo muy gratificante y está bien que no tengas que preocuparte por hacer un examen final”

En definitiva, las actividades grupales diseñadas para esta asignatura funcionan a un nivel significativo adecuado, en el futuro sería deseable seguir haciendo partícipes a los alumnos del proceso de evaluación e introducir más actividades grupales en otras partes de la asignatura.

La experiencia de trabajo en grupo llevada a cabo en las asignaturas de Cálculo y Matemáticas Asistidas por Ordenador nos confirma que se debe intentar enseñar a trabajar con otros. Así, en la medida en que nuestros alumnos se acostumbren y aprendan a trabajar en equipo, podrán resolver más eficazmente situaciones conflictivas y conducir mejor equipos de trabajo futuros.

Bibliografía

- Artzt, A.F. y Newman, C. M. (1990). Cooperative learning. *Mathematics Teacher*, 83: 448-453.
- Challis N., Cox B., Fedjki Ch., Folres R., Goldman F., Pountney D., Price L. y Stirling D. (2003). Using Assessment to Improve Learning. *Proceedings of UMTC 2003, Birmingham*.
<http://www.umtc.ac.uk/umtc03/umtc2003proca.pdf> [ref. 1 junio 2005].
- Dirección de Investigación y Desarrollo Educativo de la Vicerrectoría Académica del Instituto Tecnológico y de Estudios Superiores de Monterrey (México). *El método de proyectos como técnica didáctica*.
<http://www.sistema.itesm.mx/va/dide/inf-doc/estrategias/proyectos.html> [ref. 1 junio 2005].
- Dupuy, George. (2004). *Self & Group Assessment*. Department of Business Administration, Presbyterian College, South Carolina, Estados Unidos.
<http://web.presby.edu/writingcenter/newsletter/groupassess.html> [ref. 1 junio 2005].
- García Raffi, Luis M., Pérez Peñalver, María José, Sánchez Pérez, Enrique A y Sánchez Pérez, Juan V. (2000). *La modelización como instrumento didáctico: La asignatura Matemáticas Asistidas por Ordenador, en la E T S I C C P de la Universidad Politécnica de Valencia*. 1^{er} Congreso Internacional: Docencia Universitaria e Innovación. Universidad Politécnica de Cataluña, Barcelona.

- Gómez Urgellés, J. (coordinador). (2003). *Modelització matemàtica com a eina d'ensenyament*. Proyecto interuniversitario subvencionado por DURSI, <http://www-ma4.upc.edu/~andreu/> [ref. 2 junio 2005].
- López-Real, F. y Rita Chan, Y. (1999). Peer Assessment of a group Project in a Primary Mathematics Education Course. *Assessment & Evaluation in Higher Education*, 24 (1): 67-79.
- MacBean J., Graham T. y Sangwin C. (2003). *Guidelines for Introducing Group Work in Undergraduate Mathematics*. Higher Education Academy - Maths, Stats & OR Network. <http://mathstore.ac.uk/projects/groupwork/>, [ref. 1 junio 2005].
- Robert, H., Stewart, B., Quinney, D. and Hibberd, S. (2002). Group Projects in Mathematical Courses. *Proceedings of UMTC 2002, Birmingham*. <http://www.umtc.ac.uk/umtc2002/umtc2002proca.pdf> [ref. 1 junio 2005].
- Tinzmann M.B., Jones B.F., Fennimore T.F., Bakker J., Fine C. y Pierce J. (1990) *What Is the Collaborative Classroom?* NCREL, Oak Brook. http://www.ncrel.org/sdrs/areas/rpl_esys/collab.htm [ref. 1 junio 2005].

Apéndice A. Parrillas de evaluación para la asignatura de Cálculo

MIEMBROS DEL GRUPO

HOJA NÚMERO

--

ASISTENCIA A REUNIONES

	1 Clase	2 Clase	3 Clase	Tutoría		

CRITERIOS PARA VALORAR EL TRABAJO DE NUESTRO GRUPO

	Asistencia a reuniones (individual)
	Nivel de implicación en las tareas y discusiones del grupo (individual)
	Capacidad de organizar las tareas del grupo (grupal)
	Capacidad de escuchar a los miembros del grupo (individual)
	Capacidad de rectificar y aceptar las críticas constructivas (individual)
	Capacidad de llegar a consenso (grupal)

Parrilla 1. Utilizada en la coevaluación

PROCESO DE AUTOEVALUACIÓN DEL TRABAJO DE NUESTRO GRUPO

MIEMBROS DEL GRUPO ⇒						
A) Asistencia a reuniones (individual)						
B) Nivel de implicación en las tareas y discusiones del grupo (individual)						
C) Capacidad de escuchar a los componentes del grupo (individual)						
D) Capacidad de rectificar y aceptar las críticas constructivas (individual)						

Parrilla 2. Utilizada en la autoevaluación

MIEMBROS DEL GRUPO

HOJA NÚMERO

--

ASISTENCIA A REUNIONES

	1 Clase	2 Clase	3 Clase	Tutoría Obligatoria	Tutoría Optativa	Tutoría Optativa

PRIMERA SESIÓN	
SEGUNDA SESIÓN	
TERCERA SESIÓN	

Parrilla 3. Hoja de control

<p>¿Cómo de efectivo las ha sido vuestro grupo en realizar las tareas?</p> <p>----- Muy productivo y extraordinariamente efectivo ----- Moderadamente efectivo ----- Bastante poco efectivo porque....</p> <p>¿Cómo calificarías la experiencia de trabajar juntos?</p> <p>----- Muy buena, volvería a trabajar con este grupo de nuevo ----- Satisfactoria, hemos realizado bastante bien la tarea ----- Decepcionante porque....</p>	<p>¿Cómo evaluarías la calidad de tus contribuciones al grupo?</p> <p>---- Excelente, muy útil y constructiva ---- Bastante buena, pero no insuperable ---- Buena, pero podría haberlo hecho mejor ---- Pobre, no me he implicado suficientemente</p> <p>Señala qué cosas has aprendido tú y tu grupo de cómo se trabaja bien en grupo</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Parrilla 4. Preguntas para la evaluación final del proceso del trabajo en grupo (basado en Dupuy, 2004)

Apéndice B. Parrillas de evaluación para la asignatura de Matemáticas Asistidas por Ordenador

EVALUACIÓN TRABAJO DE MAO

Componentes del grupo:

Título del proyecto:

--

PRODUCTO

	MUY BAJO (2)	BAJO (4)	MEDIO (6)	ALTO (8)	MUY ALTO (10)
A) DISEÑO GLOBAL Recopilación de información Extensión y profundidad					
B) CONTENIDO CONCEPTUAL Planteamiento y rigor matemático Resolución e interpretación Corrección en los resultados					
C) PRESENTACIÓN Estructura y claridad de la presentación Explicaciones orales claras y precisas					
D) ACTITUD CIENTÍFICA, CREATIVIDAD					
E) CONCLUSIONES Y COMENTARIOS					

$$0.2 * A + 0.3 * B + 0.2 * C + 0.2 * D + 0.1 * E$$

PROCESO

		Miembros del grupo (distribuir 100 puntos entre los miembros) Evitar una puntuación por igual para todos				
		GRUPO (0-5)				
A) REPARTO EQUITATIVO DE TAREAS						
B) INICIATIVA						
C) CONOCIMIENTO DE LOS CONTENIDOS DEL TRABAJO						
D) GENERACIÓN DE ALTERNATIVAS A LOS PROBLEMAS						
E) NIVEL DE IMPLICACIÓN						

La parrilla del producto está basada en la propuesta por Gómez Urgellés (2003)



Relación entre autoevaluación y evaluación por profesor. Estimación del tiempo de dedicación del estudiante

*Frances Watts
Amparo García-Carbonell
Nieves Martínez Alzamora*

El presente trabajo recoge los resultados obtenidos de la investigación llevada a cabo con los estudiantes que conforman los grupos de Inglés Avanzado, en la Escuela Técnica Superior de Ingenieros de Telecomunicación de la Universidad Politécnica de Valencia, durante el curso 2004-2005. El trabajo en grupo como estrategia docente que se aplica en la asignatura propicia la efectividad de la metodología de simulación y juego y conlleva la necesidad de emplear instrumentos de evaluación diagnóstica, formativa y sumativa. El objetivo principal de este estudio es comparar la autoevaluación del alumno con la evaluación por el profesor. Un segundo objetivo es realizar la estimación del tiempo de dedicación individual no presencial invertido en la asignatura por los alumnos. Los resultados obtenidos indican que no existe una diferencia significativa entre la autoevaluación y la evaluación por el profesor, lo que podría considerarse una muestra de consolidación del sistema de evaluación implantado.

Palabras clave: *autoevaluación, evaluación por profesor, evaluación continua, trabajo en grupo, simulación y juego.*

1. Introducción

El actual marco de referencia europeo acoge un nuevo paradigma educativo que demanda no sólo un cambio en los períodos que satisfacen las titulaciones, sino también en los contenidos que constituyen los programas y su evaluación, en las estrategias docentes y en la distribución del tiempo presencial y no presencial que

hasta el momento requiere cada crédito. Los créditos en el sistema europeo de transferencia (ECTS) contemplan tanto el tiempo invertido en el proceso de instrucción como el del proceso de reflexión-asimilación, por lo que el cómputo de créditos se ha de calcular de manera distinta.

La trayectoria investigadora propia en el campo de la evaluación revela aspectos que permiten la sistematización de la evaluación en la asignatura de Inglés Avanzado y la consolidación de criterios de valoración, al tiempo que facilita presentar a los alumnos las pautas de evaluación con mayor claridad y rigor. Este estudio explora el grado de aceptación de dichas pautas y establece la relación entre la evaluación por el profesor y la autoevaluación. Un segundo objetivo trata de acotar el tiempo de dedicación individual necesario para cumplir con los requisitos que la asignatura demanda.

El trabajo sobre evaluación que aquí se presenta es la continuación de una trayectoria investigadora que tiene sus comienzos en los años 90, con la implantación de la simulación telemática como estrategia docente. En la simulación los contenidos surgen, por un lado, del escenario y la trama que la simulación aporta y, por otro, de las fases de *preparación*, *acción* y *puesta en común y reflexión* que toda simulación implica. Tanto el aspecto didáctico como de contenidos cuentan ya con unos resultados cualitativos y cuantitativos que muestran el estudio y medición de variables como, por ejemplo, efectividad del método, progreso del conocimiento o afectividad¹, entre otras.

Los estudiantes de la ETSI de Telecomunicación que cursan la asignatura de *Inglés Avanzado* han participado en diversas ediciones de simulaciones telemáticas asistidas por ordenador a gran escala, donde grupos procedentes de diferentes países y disciplinas participan en una discusión y negociación a nivel internacional, inmersos en un escenario común. La primera experiencia tiene lugar con el proyecto IDEALS², seguido de la participación en el proyecto ICONS³ y en el proyecto IDEELS⁴; todos ellos diseñados siguiendo la filosofía de un primer proyecto, el proyecto POLIS⁵.

¹ Bredemeier y Greenblat (1981); Miles et al. (1986); Wolfe (1985) (1987); Wolfe y Teach (1987); Wolfe y Roberts (1993); García Carbonell (1998); Rising (2001); Peters (2005), entre otros.

² Proyecto IDEALS: *International Dimensions in Education via Active Learning and Simulation*. Universidad de Alabama.

³ Proyecto ICONS: *International Communication and Negotiation Simulations*. Universidad de Maryland.

⁴ Proyecto IDEELS: *Intercultural Dimensions in European Education through On-line Simulation*. Universidad de Bremen (coordinadora).

⁵ Proyecto POLIS: *Policy Simulation*. Universidad de California.

Tanto ICONS como IDEALS brindan una experiencia educativa interactiva, multipuesto, que cubre una variedad de áreas de conocimiento relacionadas con asuntos de interés mundial. ICONS ofrece una serie de ejercicios de simulación que operan en un mundo real, fundamentalmente temas de política exterior, comunicación, lenguas; mientras que, IDEALS opera como un planeta ideal, que utiliza la simulación como herramienta estratégica para fomentar el entendimiento internacional y desarrollar destrezas para la comunicación intercultural.

En ambos proyectos, la efectividad de la simulación se apoya en el principio de aprendizaje basado en la experiencia. La experiencia en sí misma motiva a los estudiantes, de forma que éstos sufren un proceso de inmersión, a la vez que se responsabilizan de su propio aprendizaje (García-Carbonell y Watts, 1997). Como en la mayoría de las actividades de simulación y juego, el aprender haciendo se convierte en una realidad; o como nos recuerdan las palabras del antiguo proverbio chino: oigo y olvido, veo y recuerdo, hago y entiendo.

El componente fundamental de ambos proyectos es una simulación asistida por ordenador, con un tipo de formato nuevo: un juego asíncrono de varias semanas de duración, durante las cuales también se establece una interacción de juego síncrono. Los grupos participantes adoptan el papel de altos funcionarios, representantes de varios países involucrados en negociaciones, o que asisten a diferentes foros internacionales. Los equipos proceden de diferentes partes del mundo, pero operan bajo una misma plataforma que permite la interacción en tiempo real y diferido de todos los participantes.

El proyecto IDEELS⁶ tiene como objetivo crear, experimentar y evaluar una serie de escenarios donde estudiantes de distintas disciplinas ponen en práctica sus destrezas comunicativas, lingüísticas e interculturales, a la vez que afrontan los retos más importantes con los que Europa se enfrenta en la actualidad. Los participantes en simulaciones IDEELS también se comunican vía internet, utilizando un navegador para acceder al entorno creado, en el que pueden representar a un país, un grupo de asesores, o una empresa periodística, entre otros. En la actualidad, IDEELS cuenta con seis escenarios para Eutropía, un mundo utópico que refleja la riqueza que ofrece la diversidad cultural de Europa. Los temas que se han abordado en diferentes simulaciones han sido los derechos humanos, la inmigración, el sistema educativo, las tecnologías de la información y comunicación y el derecho de agua.

⁶ IDEELS es un proyecto Socrates/Erasmus diseñado por un grupo de profesores-investigadores procedentes de las universidades de Bremen (Alemania), Nice-Sophia Antipoli (Francia), Bergen y Nord Trondelag (Noruega) y la Universidad Politécnica de Valencia, UPV (España). <http://www.ideels.uni-bremen.de>

La participación, no sólo en el diseño sino como usuarios de IDEELS, permite presentar los primeros resultados cuantitativos en evaluación, tercer aspecto del nuevo paradigma europeo. Esta investigación, como se apunta anteriormente, comienza a mediados de los años noventa y continua con escenarios de enseñanza-aprendizaje con un enfoque más cooperativo, donde la reflexión es imprescindible, no sólo durante la adquisición de conocimientos, sino también durante la evaluación. Los resultados que aquí se cuantifican responden a un tipo de evaluación que en sí misma transcurre por las fases del círculo de aprendizaje experiencial que Kolb (1984) define; este estudio supone el comienzo de una serie de estudios que centran sus hipótesis en esta línea. El concepto circular de aprendizaje que se postula parte de una Experiencia Concreta para volver a ella, tras pasar por las fases de Observación Reflexiva, Conceptualización Abstracta y Experimentación Activa. Este enfoque no sólo constituye un importante principio en la educación experimental, sino que, la evaluación a través del portafolio también asume esos principios. Elaborar el portafolio es una experiencia concreta que obliga a reflexionar y abstraer, de ahí que constituya una experiencia de aprendizaje en sí misma. Esta reflexión tiene como finalidad mostrar la filosofía común entre el método de la simulación y su evaluación.

1.1. Trabajo en grupo

En la asignatura que nos ocupa, el trabajo en grupo es una herramienta adecuada para canalizar la metodología propuesta. Cada clase es un equipo que adopta una identidad y desempeña un papel en la simulación. Los equipos restantes están en distintas partes del mundo, pero operan bajo una misma plataforma de comunicación que les permite la interacción en tiempo real y diferido. A su vez, cada clase se divide en equipos de trabajo más pequeños a los que se les asigna unos temas y unas tareas concretas. La formación de estos equipos responde a un criterio general de paridad en el conocimiento de la lengua inglesa, de forma que el nivel intraequipos sea lo más homogéneo posible. La asignación y distribución de otros perfiles y tareas adicionales de apoyo al cabeza de grupo, de portavoz, secretario, controlador lingüístico, etc., goza de mayor libertad.

1.2. La evaluación en la asignatura de Inglés Avanzado

En un primer acercamiento a la evaluación utilizamos el concepto de De la Cruz (1999), quien plantea la evaluación como un proceso de recogida de información para utilizarla en la formulación de juicios y la toma de decisiones que son necesarias en todo proceso de aprendizaje. Es bastante común observar cómo el significado de evaluar se reduce a juzgar y calificar, es decir, asignar un valor numérico al rendimiento. En esta línea, alguna de las puntualizaciones hechas por la propia autora nos servirán para matizar algunos aspectos.

- La evaluación es un proceso, con diversidad de momentos y tareas, no es una actividad aislada y con fin en sí misma.
- El objetivo de la evaluación es tomar decisiones bien informadas sobre todos los componentes del proceso de enseñanza-aprendizaje, sus efectos o resultados, sus procedimientos y sus protagonistas. El tipo de decisiones a tomar, como objetivo final del proceso de evaluación, condiciona el tipo de información a recoger, las circunstancias de su obtención y los procedimientos o técnicas a utilizar (De la Cruz 1999: 22).

La evaluación del aprendizaje de los alumnos en la asignatura de Inglés Avanzado es una evaluación continua; es decir, a lo largo del curso se produce una evaluación formativa y al término del curso una prueba final, evaluación sumativa. La evaluación formativa sirve para comprobar el grado de adquisición del contenido por el alumno y, a la vez, le proporciona información de retorno (*feedback*). La evaluación sumativa tiene como finalidad comprobar si los aprendices han alcanzado los objetivos propuestos y sirve para la toma de decisiones en cuanto a la promoción académica. Las pruebas de evaluación sumativa son relativamente sencillas de confeccionar porque se guían enteramente por el programa y miden los contenidos que aporta la información entrante, el *input*; por ello, su validez está relacionada con el contenido y la metodología del programa.

La evaluación continua que se plantea requiere de información adecuada sobre la marcha del proceso de aprendizaje. Precisa de una comunicación fluida entre el profesor y el alumno, para poder ratificar lo que va bien y rectificar lo que sea preciso. El alumno, por su parte, ha de llegar a conocer su situación personal (lo que sabe, lo que duda, lo que ignora, lo que confunde, etc.) a través de distintas actividades, correcciones, iniciativas, nivel de participación, del trabajo tanto individual como en grupo, y de las observaciones e indicaciones constantes del profesor. Por ello, la evaluación continua que se propone aglutina una evaluación orientativa, formativa y sumativa.

En primer lugar, la evaluación orientativa inicial que se lleva a cabo determina los distintos niveles de conocimiento de la lengua de los alumnos que integran el grupo y procura una distribución equitativa para el trabajo en equipo. Para formar los equipos se realiza una prueba inicial rápida de respuesta múltiple que principalmente mide aspectos gramaticales. En una primera etapa se utilizó el Oxford Placement Test de 1985, sustancialmente modificado a raíz de los resultados del análisis de items que mostraron lo inadecuado de muchas preguntas para los alumnos (Watts y García Carbonell 1996). Actualmente se utiliza el Oxford Placement Test de 2001 que produce resultados más fiables, a pesar de las carencias que muestran los análisis de items realizados anualmente. Se ha observado la eficacia del criterio de distribuir a los alumnos en los equipos equitativamente por nivel, de

modo que al frente de cada equipo de trabajo quedan como redactores-jefe los alumnos con el nivel de gramática más alto.

En segundo lugar, la evaluación formativa arranca en la primera fase de la simulación con el estudio de documentación en relación con el tema de la simulación. El escenario, las identidades, el perfil y la situación de partida propia y de los otros participantes, son algunos ejemplos. La tarea de preparar dos documentos que representen la posición inicial del equipo y las estrategias de futuro obliga a los estudiantes a reunirse con los compañeros de equipo fuera de las sesiones regladas para poner en común ideas e información. Una matriz SWOT, siglas de las palabras en inglés *strengths*, *weaknesses*, *opportunities* y *threats*, permite reflexionar sobre los puntos fuertes, los puntos débiles, oportunidades y amenazas. De igual modo, permite valorar y redactar los documentos que contienen, por un lado, la posición política para conocimiento público y, por otro, las estrategias a seguir.

En las sesiones presenciales, se intercalan simulaciones situacionales cortas para familiarizar a los alumnos con la dinámica del trabajo en equipo. Después de cada simulación corta se hace una puesta en común y evaluación con todo el grupo, guiada por preguntas del profesor-facilitador o por una herramienta cognitiva del tipo propuesto por Edward de Bono (1996). De este modo, los alumnos reflexionan y evalúan lo que acaban de aprender en el ejercicio y consolidan lo aprendido. Si se recuerda que el aprendizaje se produce de forma circular y en diferentes etapas, estas sesiones evaluativas forman parte de la fase de reflexión. El profesor-facilitador tiene además la ocasión de observar si los alumnos han entendido la nueva metodología y el grado de participación y de compromiso entre compañeros de equipo. Las sesiones evaluativas se suceden a lo largo de las 4-5 semanas más intensas del curso, donde tienen lugar además, aclaraciones generales, individuales, cambio de equipo, etc.

La evaluación formativa también incluye la realización de una redacción en un tiempo asignado, cuyo propósito es ayudar al alumno a plasmar por escrito un destilado de los conocimientos adquiridos hasta el momento, tanto del contenido del tema de la simulación como su habilidad en el manejo de técnicas de redacción. Esta prueba puntúa para la calificación final.

La evaluación sumativa final permite al profesor-facilitador medir y juzgar el aprendizaje global del alumno. El elemento primordial que debe servir de base en toda evaluación final es el grado de consecución de los objetivos de la asignatura y los contenidos desarrollados en el programa (Contreras 1990: 88).

En este caso, los alumnos elaboran un portafolio que refleja la experiencia de su participación en la simulación. Éste contiene información de las tres fases del proceso, se reflexiona sobre la preparación de la simulación, las clases magistrales, los debates, la acción acontecida entre equipos durante la segunda fase de negociación,

la puesta en común y evaluación de la simulación. El portafolio se utiliza como vehículo para que los estudiantes argumenten y discutan ideas conjuntamente y ofrece testimonio de la comprensión y capacidad de los alumnos para aplicar lo que han comprendido. Adams y Hamm (1998: 31) apuntan que con el portafolio la atención se centra en la actuación y aplicación de conocimientos y no en el conocimiento por el conocimiento.

El proceso de reflexión que se realiza durante la elaboración del portafolio ayuda a los estudiantes a ser conscientes de su reciente historia de aprendizaje y de su habilidad para razonar de forma más científica. El estudiante se involucra de tal forma en la tarea de valoración que la barrera entre él y su evaluación disminuye. A través de un análisis crítico de su propio trabajo –y del trabajo en equipo– los alumnos llegan a conocer diferentes formas de resolver problemas. A la vez que llegan a conocer mejor lo que saben del contenido, su capacidad en el manejo de la lengua y su forma de pensar se hacen más expertos en la evaluación de su propio trabajo.

El portafolio contiene los materiales de trabajo y nos muestra los conocimientos y la actuación de su autor. Esto permite al profesor-facilitador conocer con más exactitud el trabajo del estudiante, su desarrollo personal del conocimiento y su madurez. A grandes rasgos, se puede decir que los alumnos en el diseño del portafolio han de cuidar especialmente:

- qué se ha valorado,
- el diseño del portafolio,
- la adecuación de los contenidos,
- la corrección lingüística,
- el posible lector (del documento escrito) y
- la audiencia (en la defensa oral del trabajo).

Una actividad adicional de reflexión y evaluación es la presentación y defensa oral de los portafolios, donde se aprecia el nivel de competencia oral adquirido en la lengua. Durante la simulación, los estudiantes practican las destrezas orales en su interacción con el equipo y el grupo en general. Si se les guía para que interactúen de forma eficaz, son capaces de desarrollar estrategias comunicativas que pueden aplicar en escenarios de la vida real. En la simulación interaccionan con perfiles de la vida real profesional, donde, por ejemplo, han de explicar algún tipo de investigación llevado a cabo, debatir sobre la política de las compañías, analizar problemas o dar instrucciones de trabajo, hacer presentaciones orales o redactar informes.

La presentación oral de sus portafolios en presencia del grupo es el último paso en el proceso de la simulación. Los alumnos conocen los criterios de evaluación que se aplican en la valoración del discurso oral y, para la presentación, seleccionan los aspectos que consideran más relevantes, lo que produce un perfil más completo de su aprendizaje (García-Carbonell, Watts y Rising 1998: 335). (Ver Apéndice D).

Los distintos pasos que se han seguido en la evaluación de la asignatura tratan de conseguir unas cualidades propias de las pruebas de lenguas que, en opinión de Bachman y Palmer (1996), deben tener autenticidad, interactividad, impacto y practicabilidad. No se debe potenciar, en su opinión, una única cualidad sobre otra, sino propiciar un efecto combinado, en aras de conseguir la máxima utilidad de la evaluación.

En resumen, la evaluación en la asignatura de Inglés Avanzado se interpreta como parte del aprendizaje. Inicialmente, los alumnos averiguan su nivel. A lo largo del curso, se hacen puestas en común para valorar y evaluar el proceso. Al final, los alumnos sintetizan los contenidos, lo acontecido y aprendido, defienden su visión de los hechos y terminan plasmando por escrito toda la experiencia; todo basado en la reflexión y análisis inducido. El profesor, por su parte, se vale de una prueba objetiva de elección múltiple para la prueba diagnóstica inicial, de cuestionarios y herramientas cognitivas para la evaluación del proceso y de criterios analíticos para la evaluación del producto. La variedad en los instrumentos y medidas de evaluación favorece el aprendizaje a la vez que contribuye a lograr la validez y fiabilidad que debe garantizar toda evaluación justa.

1.3. Investigación previa sobre la evaluación en la asignatura

La evaluación en la asignatura de Inglés Avanzado del plan de estudios de la ETSI de Telecomunicación ha sido a lo largo de los años objeto de diversos trabajos, que se han solapado con los estudios sobre la efectividad de la metodología utilizada. Los primeros esfuerzos se centran en perfeccionar la prueba diagnóstica, ya que el test comercial no se ajustaba a la realidad de los alumnos de Inglés Avanzado; no alcanzaba el nivel de dificultad adecuado y el índice de discriminación no era correcto (Watts y García Carbonell, 1996). La validación de criterios de corrección de textos escritos (Watts 1997; Watts y García Carbonell, 1998 y 1999) fue punto de partida para la concreción de criterios para la evaluación del portafolio (García Carbonell, Watts y Rising, 1998), que a su vez sirvieron el propósito de formar parte del estudio de la efectividad de la simulación telemática en el aprendizaje del inglés técnico (García Carbonell, 1998). Los criterios para evaluar las presentaciones orales fueron objeto de otro estudio (García Carbonell, Gotor, Montero, Rising y Watts, 2001). Finalmente, un trabajo sobre el estado de la

cuestión del uso de la simulación y juego en la adquisición de la competencia comunicativa en lenguas extranjeras (García Carbonell, Rising, Montero y Watts 2001) recoge toda la investigación llevada a cabo hasta la fecha.

1.4. Tiempo de dedicación del alumno

Como se ha mencionado al principio de la introducción, en el sistema europeo de transferencia de créditos, no sólo se contempla el tiempo invertido en el proceso de instrucción sino también el del proceso de reflexión-asimilación, es decir, tanto el tiempo que el alumno está presente en clase como el tiempo que utiliza fuera del aula para preparar y superar la asignatura. Este tiempo no presencial se puede invertir en diversas actividades tales como tutorías, consultas, preparación de trabajos y exámenes, estudio personal, entre otros. Los créditos ECTS pretenden contemplar la actividad global de aprendizaje del estudiante. Lavigne (2003) apunta tres métodos básicos de adaptación al sistema europeo: el reparto de los créditos de los que se dispone entre las disciplinas o módulos que componen una titulación (impositivo o analítico), el cálculo a partir del tiempo de trabajo que emplea el estudiante (compositivo o sintético) y la estimación del tiempo que implica la consecución de las competencias definidas para una disciplina (competencial). A nivel institucional, quizás sea aconsejable una combinación de los tres métodos, aunque se puede avanzar en la preparación para el cambio mediante la exploración de la segunda modalidad, el control del tiempo que emplea el estudiante en conseguir los objetivos previstos en las asignaturas.

1.5. Objetivos del presente estudio

Si se tiene en cuenta la investigación previa sobre la evaluación en la asignatura y las mejoras implementadas, este estudio indaga como primer objetivo el grado de aceptación del sistema de evaluación por parte del alumno, comparando la autoevaluación del alumno con la evaluación por el profesor. Un segundo objetivo es obtener una estimación del tiempo no presencial invertido en la asignatura, para poder sumarlo al tiempo presencial y obtener el cómputo global de tiempo, con el fin de calcular con mayor exactitud los créditos transferibles en el marco del espacio europeo de educación superior.

2. Materiales y método

2.1. Materiales

En este trabajo se consideran fundamentalmente tres variables: evaluación por el profesor, autoevaluación y tiempo de dedicación individual no presencial. En el tratamiento de la primera variable, *calificación del profesor*, se trabaja en consideración

a cuatro factores: una redacción cronometrada (5% de la calificación final), el portafolio escrito final (80%), la presentación oral del portafolio (10%) y nivel de participación en la asignatura (5%). En la redacción cronometrada se utiliza un formulario de encabezado y formato único para todos los estudiantes; para su evaluación se utiliza la parrilla de criterios definidos (consultar Apéndice B). El portafolio escrito, a grandes rasgos, reúne los documentos de estudio relacionados con el tema de la simulación y los documentos producidos, comenta las negociaciones entre equipos participantes, recopila los resultados de las teleconferencias, aporta la opinión personal, presenta conclusiones e introduce propuestas de futuro. Tanto para la valoración del portafolio como para la presentación oral del mismo se utiliza una parrilla de criterios definidos (consultar Apéndices C y D). Otro de los materiales que se utiliza es la lista de preguntas que guían la reflexión que se ha de plasmar en el portafolio escrito (consultar Apéndice A). En la variable *tiempo individual no presencial* se utilizan plantillas de tiempos que recogen el cómputo de tiempo individual semanal de dedicación a la asignatura de cada estudiante. El análisis estadístico se lleva a cabo a través de la aplicación SPSS para Windows (Versión 12.0).

2.2. Método

El número total de alumnos de Inglés Avanzado del año académico 2004-05 que participa en el estudio es de 37. En la primera fase de la simulación se obtiene los resultados de la prueba cronometrada. Los alumnos, en un espacio de tiempo de 30 minutos, redactan sobre el tema central que se plantea en la simulación. La segunda y tercera fase de la simulación ofrecen las calificaciones del portafolio escrito, de la defensa oral y de la participación en clase. Durante las tres fases se va obteniendo el cómputo del tiempo individual no presencial, con un total de 26 alumnos (70'3%). Finalmente, del total de alumnos participantes, un número de 28 (75'7%) se autoevalúan una vez finalizado el curso y el portafolio. En definitiva, la muestra se obtiene en los distintos estadios del curso. Tras la recopilación de datos se procede al análisis estadístico e interpretación de resultados.

Se verifica normalidad en las variables *calificación del profesor*, *autoevaluación* y *tiempo individual no presencial* mediante el test de Kolgomorov-Smirnov, no encontrando evidencia para rechazar el modelo normal en ninguna de ellas ($\alpha = 0'05$).

Con objeto de comparar la autoevaluación y la evaluación realizada por el profesor, se utiliza la prueba t para muestras relacionadas, calculando las diferencias entre los valores de las dos variables para cada alumno y contrastando si su media difiere significativamente de cero para un nivel de significación $\alpha = 0'05$.

Con objeto de estimar el tiempo de dedicación individual necesario para preparar la asignatura, se obtiene un intervalo de confianza al 95% para el tiempo no presencial dedicado por el alumno a esta asignatura. No se tiene en cuenta el tiempo invertido en realizar el portafolio, ya que no se dispone de suficiente información al respecto.

Para analizar el valor predictivo que las variables tiempo individual de dedicación no presencial y autoevaluación pueden tener sobre la evaluación del profesor se ajusta un modelo de regresión a los datos. La selección de variables se realiza mediante el método *stepwise*, siendo el nivel de significación para la incorporación de nuevas variables $\alpha = 0.05$ y el nivel de significación para la eliminación de variables $\alpha = 0.10$.

3. Análisis de resultados

3.1. Relación entre autoevaluación y evaluación por profesor

Para los alumnos de Inglés Avanzado de la ETSI de Telecomunicación, el participar en una simulación a gran escala es una experiencia de inmersión en la lengua que añade una nueva dimensión a la enseñanza en el marco universitario. La organización de la clase en equipos de trabajo, a los que se les asigna unos temas y unas tareas concretas, aporta a los alumnos una valiosa experiencia en la dinámica de trabajo en grupo y las consiguientes destrezas comunicativas de definición de estrategias, resolución de problemas o toma de decisiones. Se trata de una metodología ya consolidada y su eficacia en la mejora del aprendizaje de los alumnos en el contexto concreto de la UPV está demostrada en investigaciones previas. Uno de los aspectos pendiente de estudio es el que actualmente se presenta: el grado de aceptación por parte de los alumnos del sistema de evaluación que se ha ido imponiendo a lo largo de los años.

Los resultados de este estudio (consultar Tabla 1) señalan una media muestral de las evaluaciones realizadas por el profesor (7.82) ligeramente superior a la media muestral de las evaluaciones realizadas por los alumnos (7.66). El test t de *Student* para muestras relacionadas (consultar Tabla 2) indica que esta diferencia no es significativa (p-valor = 0.26). Es decir, los resultados indican que los alumnos tienen un conocimiento muy aproximado de la calificación que pueden esperar al terminar la asignatura.

	N	Mínimo	Máximo	Media	Desv. típica	Recorrido Inter.
Evaluación profesor	37	5,0	10	7,82	1,47	2'30
Autoevaluación	28	6,0	10	7,66	1,06	1'65
Tiempo estudio no presencial	26	18	63	40,69	14,68	27'63

Tabla 1. Estadísticos descriptivos de la comparación entre evaluaciones

Prueba de muestras relacionadas	Diferencias relacionadas				t	gl	Sig. (bilateral)	
	Medida	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
				Inferior				Superior
Par 1 Evaluación profesor - Autoevaluación	,24821	1,15153	,24762	-,19830	,69473	1,141	27	,264

Tabla 2. Prueba t de *Student* de comparación entre evaluaciones

La dispersión es ligeramente superior para las evaluaciones realizadas por el profesor, siendo el rango de valores 5-10 para la evaluación del profesor y 6-10 para la autoevaluación. El recorrido intercuartílico es 2'30 para la evaluación del profesor y 1'65 para la autoevaluación (ver Figura 1).

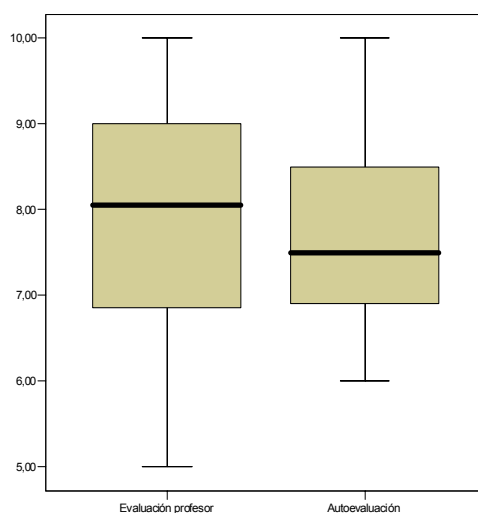


Figura 1. Evaluación por el profesor y autoevaluación

Una mayor dispersión en las calificaciones del profesor evidencia que se evitan errores atribuibles a la variable juez, como la gravitación hacia el punto medio de una escala y el efecto halo. Estos resultados se podrían considerar como muestra de la consolidación del sistema de evaluación implantado.

3.2. Tiempo individual no presencial

El tiempo medio no presencial dedicado por los alumnos para la preparación de la asignatura, a excepción del invertido en la elaboración del portafolio, es de 40'69 horas con un error estandar de 2'87 horas (I.C. 95% = (34'76, 46'62)). El tiempo mínimo es de 18 horas y el tiempo máximo es de 63. El recorrido intercuartílico es 27'63 horas (ver Figura 2).

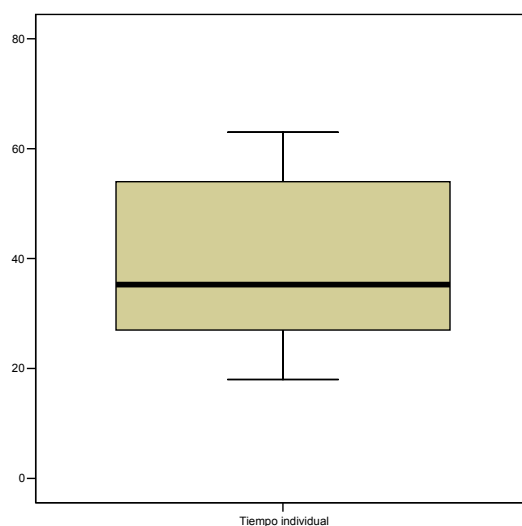


Figura 2. Tiempo individual no presencial (no incluye tiempo de elaboración del portafolio)

La variable tiempo no presencial no se incorpora como variable explicativa al modelo de regresión por el método *stepwise*, lo cual indica que en la asignatura de Inglés Avanzado no existe una relación significativa entre el tiempo de dedicación individual y calificación final. La autoevaluación se incorpora al modelo con un coeficiente que no difiere significativamente de uno, confirmando el hecho de que no existe una diferencia significativa entre la autoevaluación y la evaluación del profesor. El coeficiente R^2 del modelo ajustado es 0'41 (consultar Tabla 3 y ver Figura 3).

Modelo	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Intervalo de confianza para B al 95%	
	B	Error típ	Beta			Límite inferior	Límite superior
1 Autoevaluación	1,030	,028	,990	36,488	,000	,972	1,088

a. Variable dependiente: evaluación profesor

b. Regresión lineal a través del origen

Tabla 3. Ajuste de un modelo de regresión para la variable evaluación del profesor

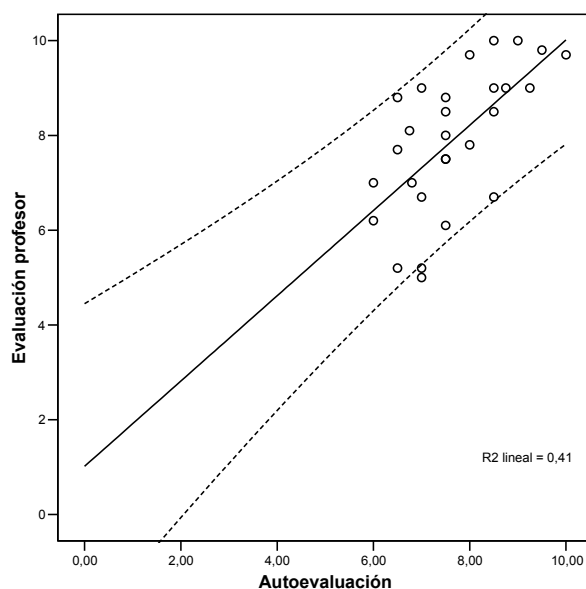


Figura 3. Diagrama de dispersión: evaluación profesor, autoevaluación

En este estudio, se ha hecho un primer intento de calcular el tiempo que los alumnos utilizan para preparar la asignatura durante el curso. En las trece semanas que dura el curso está previsto que los alumnos necesitan 31 horas de dedicación individual de estudio (26 para preparar las clases y 5 para preparar la presentación y defensa oral del portafolio). El tiempo necesario previsto para elaborar el portafolio es de 40 horas. En el tiempo individual no presencial, no se ha cuantificado el tiempo utilizado en las sesiones semanales de clase, las teleconferencias, las reuniones de los grupos de trabajo fuera del horario de clase, ni la asistencia a tutorías o sesiones especiales. Se trata de tiempos de fácil registro que deberán sumarse junto al tiempo de dedicación individual no presencial para la adaptación a ECTS.

La media hallada de los datos obtenidos del registro semanal de tiempo individual no presencial (40'69 horas) es superior al tiempo indicativo propuesto (31 horas). La falta de datos objetivos sobre la dedicación a la preparación del portafolio ha imposibilitado el estudio de la relación entre las variables *tiempo de dedicación no presencial* y *calificación del profesor*. Para completar la investigación y, por tanto, continuar el estudio de la carga de trabajo individual del alumno, será necesario pedir a los estudiantes que incluyan el cómputo en el mismo portafolio. De este modo, en el próximo análisis se contará con el tiempo individual dedicado tanto a la preparación de las clases como a la elaboración del portafolio.

5. Conclusión

La evaluación de una asignatura ha de adecuarse no sólo a los contenidos sino a la metodología utilizada, que a su vez ha de alcanzar los objetivos de aprendizaje planteados. En el caso de la asignatura de Inglés Avanzado, se utiliza una metodología activa como es la simulación telemática para lograr un amplio abanico de destrezas lingüísticas (comprensión y expresión oral y escrita) y destrezas comunicativas (definición de estrategias, toma de decisiones o resolución de conflictos, entre otras) mediante el trabajo en grupo. El sistema de evaluación que se plantea utiliza varios instrumentos de medición (la redacción cronometrada, el portafolio escrito y la presentación y defensa oral del portafolio) que requieren del alumno que produzca muestras de una serie de destrezas que incrementan el aprendizaje cognitivo global, puesto que la evaluación constituye un aprendizaje en sí misma.

La pluralidad de los instrumentos, en principio, satisface los requisitos de validez de la evaluación, es decir, se mide lo que se pretende evaluar. De igual modo, los criterios de corrección abarcan aspectos distintos de una comunicación fluida y precisa, a la vez que son claros y fáciles de aplicar, por lo que los resultados de la evaluación serán fiables y repetibles. Los resultados del presente estudio, en el que se han comparado las variables de la autoevaluación y nota del profesor, parecen indicar que los alumnos comprenden y aceptan el sistema de evaluación propuesto, lo que podría considerarse una muestra de la consolidación del sistema de evaluación implantado.

No obstante, en el baremo que se aplica en la calificación final, queda por redefinir la participación, rasgo del baremo que todavía no se ha estudiado en profundidad. En la actualidad se ajusta únicamente a la variable asistencia a clase, que tiene la ventaja de ser un parámetro totalmente objetivo. Cabe explorar la integración de la coevaluación o evaluación entre pares para valorar la participación de los alumnos. Dicha integración será objeto de estudio en una investigación futura. Por último, una fase final de este estudio será incorporar tanto el tiempo dedicado a la preparación de las clases como el tiempo dedicado a la elaboración del portafolio.

Bibliografía

- Adams, D. y Hamm, M. (1998). *Collaborative Inquiry in Science, Math and Technology*. Portsmouth, NH: Heinemann.
- Bachman, L. y Palmer, A. (1996) *Language testing in practice*. Oxford: Oxford University Press.
- Bredemeier, M. E. y Greenblat, C. S. (1981). The educational effectiveness of simulation games: A synthesis of findings. *Simulation & Games* 12(3): 307-332.
- Contreras Muñoz, E. (1990). *El profesor universitario y la evaluación de los alumnos*. Madrid: Instituto de Ciencias de la Educación de la Universidad Politécnica de Madrid.
- De Bono, E. (1996). *Serious creativity*. Woombye (AU): Six Hat Systems.
- De la Cruz, A. (1999). Evaluación del conocimiento y su adquisición. Apuntes del Curso sobre Evaluación del Aprendizaje, Instituto de Ciencias de la Educación, Universidad Politécnica de Valencia, junio 1999.
- Kolb, D. (1984). *Experiential learning: Experience as a source of learning*. Englewood Cliffs, NJ: Prentice-Hall.
- García Carbonell, A. (1998). *Efectividad de la simulación telemática en el aprendizaje del inglés técnico*. Tesis doctoral no publicada. Valencia: Universidad de Valencia.
- García Carbonell, A., Gotor, A., Montero, B., Rising, B. y Watts, F. (2001). Oral presentation assessment of simulation portafolios. En Musci E. (ed.). *On the edge of the millennium: A new foundation for gaming simulation*. Bari: Edizioni B.A. Graphis, 23-31.
- García Carbonell, A. Rising, B., Montero, B. y Watts, F. (2001). Simulation/gaming and the acquisition of communicative competence in another language. *Simulation and Gaming: An International Journal of Theory, Practice and Research*. (4)481-491.
- García Carbonell, A. y Watts, F. (1996). Telematic Simulation and Language Learning. En García Carbonell, A. y Watts, F. (eds.) *Simulation Now! Simulación ¡Ya!* Valencia: Diputació de València, 585-595.
- García Carbonell, A. y Watts, F. (1997). Simulación telemática y aprendizaje de idiomas: principios, integración curricular y didáctica. En Cuellar y Castellvi, J. (eds.) *Las nuevas tecnologías integradas en la programación didáctica de lenguas extranjeras*. Valencia: Universitat de València, 193-204.

- García Carbonell, A., Watts, F. y Rising, B. (1998). Portfolio assessment in simulation for language learning. En Geurts, J. y Joldersma, C. (eds.) *Gaming/simulation for policy development and organizational change*. Tilburg: Tilburg University Press, 333-348.
- Lavigne, R. (2003). *Los créditos ECTS y los métodos de asignación de créditos*. <http://www.uv.es/oce> [ref. 13 marzo 2006].
- Matthews, C., y Marino, J. (1990). *Professional interactions: Oral communication skills in science, technology, and medicine*. London: Prentice Hall.
- Miles, W. G., Biggs, W. D. y Shubert, J. N. (1986). Student perceptions of skill acquisition through cases and general management simulation. *Simulation and Games*, 17(1): 7-24.
- Peters, V. (2005). *The evaluation of a discipline: An approach for evaluating gaming simulation*. Ponencia presentada en el XXXVI congreso Internacional Simulation and Gaming Association (ISAGA). Atlanta, USA.
- Rising, B. (2001). *La eficacia didáctica de los juegos de simulación por ordenador en el aprendizaje del inglés como lengua extranjera en alumnos de derecho, económicas e ingeniería*. Tesis doctoral no publicada. Madrid: Universidad Pontificia Comillas de Madrid.
- Watts, F. (1997). *Reliability in English language testing: Rater agreement in the Spanish university access examination*. Tesis doctoral no publicada. Valencia: Universidad de Valencia.
- Watts, F. y García Carbonell, A. (1996). A statistical study of the modified Oxford Placement Tests. En Piqué, J. y Andreu, V. (eds.) *English in Specific Settings*. Valencia: NAU Llibres, 165-170.
- Watts, F. y García Carbonell, A. (1998). Rater agreement in English language assessment in the Spanish university access examination battery. *Language Testing Update*, 23, 43.
- Watts, F. y García Carbonell, A. (1999). Control de la calidad en la calificación de la prueba de inglés de Selectividad. *Aula Abierta*, 73, 173-190.
- Wolfe, J. (1985). The teaching effectiveness of games in collegiate business courses. A 1973-1983 Update. *Simulation and Games*, 16(3): 251-288.
- Wolfe, J. y Roberts, C. (1986). The external validity of business management games. *Simulation and Games*, 17, 1: 43-54.
- Wolfe, J. y Teach, R. (1987). Three downloaded mainframe business games: A review. *Academy of Management Review*, 12: 181-192.

Apéndice A. Preguntas para la puesta en común y evaluación

(Adaptado del manual del facilitador de la simulación telemática Internacional *Communication and Negotiations Simulations*, "ICONS debriefing questions", <http://www.icons.umd.edu/>)

Al concluir el ejercicio, se entra de un periodo fundamental del proceso de la simulación. Durante la última fase de la simulación, la puesta en común y evaluación, al alumno se le extrae de su mundo simulado y se le exige centrarse de nuevo en la realidad. Las siguientes preguntas son algunas de las que se pueden formular en el debate de clase.

1. Muchos alumnos sienten la necesidad de hablar del proceso en sí mismo. Han pasado por un periodo intenso de actividad; necesitan "soltar la tensión". ¿Cuáles han sido los éxitos? ¿Los fracasos? ¿Frustraciones? ¿Aspectos favoritos? ¿Lo que menos gustó?
2. Hablar sobre el papel y actuación de los otros equipos. ¿Por qué se comportaron los equipos del modo que lo hicieron? (ejemplos: ¿personalidad? ¿falta de información? ¿interpretación precisa del carácter de un país?) ¿Cómo hubiera reaccionado el mundo real a sus propuestas? ¿Cuántas características de los otros equipos eran realistas? ¿En qué situaciones no actuaron acorde a su papel? ¿Cómo afectó esto a la simulación?
3. ¿Cuánto reflejó la simulación la realidad? ¿En qué aspectos reflejó la realidad? ¿En qué aspectos no reflejó la realidad.
4. ¿Cuáles fueron los objetivos de política exterior de tu país? ¿Cambiaron tus objetivos durante la simulación?
5. ¿Qué destrezas o estrategias de negociación utilizó tu equipo para lograr sus objetivos? ¿Crees que fueron efectivas? ¿Qué ejemplos hay en el mundo real de estos tipos de estrategias (tanto desde tu experiencia personal como en las relaciones internacionales)?
6. ¿Qué tipos de estrategias de negociación utilizaron los otros equipos? ¿Qué negociaciones fueron efectivas? ¿En qué temas estaban dispuestos los equipos a entenderse? ¿En qué temas no lo estaban?
7. ¿Quiénes fueron vuestros aliados durante la simulación? ¿Cómo se identificó vuestro país con vuestros aliados en el mundo real? ¿Trabajaron los otros países con sus aliados en el mundo real?
8. ¿Cómo hubieran reaccionado los ciudadanos de vuestro país a vuestros objetivos tus decisiones y vuestras propuestas? ¿Cómo influyen los ciudadanos de un país en su política exterior?
9. ¿Cómo afectó la limitación de tiempo a las negociaciones? ¿Existen temas en los que el grupo podría haber logrado un acuerdo si hubiera habido más tiempo en la simulación? Sobre estos temas ¿qué cuestiones quedan por abordar?

10. Comentar el papel del ordenador en la simulación. ¿Cómo obstaculizó o facilitó las negociaciones?
11. ¿Cómo funcionó tu grupo como equipo? ¿Hubo conflictos en el proceso de toma de decisiones? ¿Cómo se compara esto con la toma de decisiones en el mundo real?
12. Comentar sobre el flujo de información. ¿Fue difícil seguir las negociaciones? ¿Cómo distinguían los alumnos entre los mensajes más importantes y los menos importantes? ¿Se daba prioridad o se ignoraba determinadas comunicaciones? ¿Cómo se compara esta situación con el flujo de información en el mundo real?

Apéndice B. Criterios de evaluación para la redacción cronometrada

(traducido de la parrilla de criterios en inglés que se utiliza en clase)

	0'5	0'4	0'3	0'2	0'1
Desarrollo del tema	El tema está completamente desarrollado y con riqueza.	El tema está completamente desarrollado.	El desarrollo del tema está presente pero limitado por estar incompleto, por falta de claridad o por falta de enfoque.	El desarrollo del tema está presente pero restringido, incompleto o poco claro.	El tema está presente pero sin desarrollar.
Organización y conexión de ideas	La organización y conexión son apropiadas.	Organización controlada, conexión con pocos problemas.	La organización está parcialmente controlada; la conexión a veces está ausente o es parcial	La organización está poco controlada. Hay poca o ninguna conexión aparente.	No hay ninguna organización ni conexión aparente.
Variedad y precisión en la gramática y el vocabulario	Hay una amplia gama de estructuras sintácticas, control morfológico y vocabulario apropiado.	Se usa sintaxis simple y compleja. La morfología es casi siempre precisa.	Se usa sintaxis simple y compleja pero con errores, o restringida. El control morfológico es inconstante. El vocabulario es suficiente, aunque a veces inapropiado.	Se usa sintaxis simple pero con muchos errores; no se controla la sintaxis compleja. El control morfológico es muy limitado. Vocabulario escaso y sencillo, se aproxima al significado, a menudo es inapropiado.	Control sintáctico y morfológico muy limitado. Vocabulario restringido y repetitivo.

Apéndice C. Criterios de evaluación del portafolio

(traducido de la parrilla de criterios en inglés que se utiliza en clase)

	9+	8	7	6	5
Desarrollo del tema	El tema está completamente desarrollado y con riqueza.	El tema está completamente desarrollado.	El desarrollo del tema está presente pero incompleto, por falta de claridad o por falta de enfoque.	El desarrollo del tema está presente pero restringido, incompleto o poco claro.	El tema está presente pero sin desarrollar.
Organización y conexión de ideas	La organización y conexión son apropiadas.	Organización controlada, conexión con pocos problemas.	La organización está parcialmente controlada; la conexión a veces está ausente o es parcial.	La organización está poco controlada. Hay poca o ninguna conexión aparente.	No hay ninguna organización ni conexión aparente.
Variedad y precisión en la gramática y el vocabulario	Hay una amplia gama de estructuras sintácticas, control morfológico y vocabulario apropiado.	Se usa sintaxis simple y compleja. La morfología es casi siempre precisa.	Se usa sintaxis simple y compleja pero con errores, o restringida. El control morfológico es inconstante. El vocabulario es suficiente, aunque a veces inapropiado.	Se usa sintaxis simple pero con muchos errores; no se controla la sintaxis compleja. El control morfológico es muy limitado. Vocabulario escaso y sencillo, se aproxima al significado, a menudo es inapropiado.	Control sintáctico y morfológico muy limitado. Vocabulario restringido y repetitivo.
Formato, ortografía, puntuación, longitud, citas	El formato es impecable y atractivo. La ortografía y puntuación están libres de errores. La longitud es apropiada. Se citan las fuentes correctamente.	El formato es correcto. Los errores de ortografía y puntuación no distraen. Se citan las fuentes correctamente.	El formato, ortografía y puntuación a veces distraen. Las fuentes en ocasiones no se citan como es debido.	Los errores de ortografía y puntuación distraen a menudo. Las fuentes con frecuencia no se citan como es debido.	Los errores de ortografía y puntuación distraen mucho.

(Las calificaciones se convertirán a una escala de 8/10)

Apéndice D. Evaluación oral

Criterios para la evaluación de la presentación oral del portafolio

2 = excelente 1 = satisfactorio 0 = necesita mejorar

(Las calificaciones parciales se sumarán y se convertirán a escala de 1'5)

I. Discurso

- A. _____ Discurso natural– sin leer o estar memorizado.
- B. _____ Velocidad al hablar– ni demasiado rápido ni demasiado lento.
- C. _____ Movimientos corporales / de postura – ningún gesto que distraiga.

II. Contenido

- A. _____ El tema es adecuado para el tiempo disponible – ni demasiado limitado ni general.
- B. _____ El tema está desarrollado con detalles, datos y ejemplos relevantes que proporcionan apoyo a la idea central.
- C. _____ La presentación cumple los requisitos de tiempo – ni demasiado larga ni demasiado corta.

III. Organización

- A. _____ Introducción.
- B. _____ Uso de mecanismos de transición.
- C. _____ Puntos principales expuestos con claridad.
- D. _____ Desarrollo de ideas lógico y fácil de seguir.
- E. _____ Conclusión.

IV. Uso de la Lengua

- A. _____ Vocabulario adecuado a los oyentes – explicación de palabras difíciles o técnicas.
- B. _____ Pronunciación / Entonación.
- C. _____ Uso de una variedad de expresiones – ausencia de jerga inapropiada o informal.
- D. _____ Control efectivo de la gramática / vocabulario – precisión.

Observaciones:

(Adaptado de los criterios de Matthews y Marino, 1990)

Investigación sobre la evaluación en un curso universitario de química general

Juan Llorens

En este trabajo se describen diferentes innovaciones didácticas relacionadas con las estrategias y métodos de evaluación en el contexto de un curso de química general dentro de los estudios de Ingeniería Técnica Agrícola. En las propuestas presentadas se concede una atención preferente a la evaluación formativa y a su influencia en el rendimiento académico. Asimismo, desde una perspectiva teórica basada en el concepto de “zona de desarrollo próximo” de Vygotsky, se estudia la relación entre el rendimiento obtenido en tareas de evaluación realizadas en grupo, respecto de sus resultados individuales. De igual modo, se analiza en qué medida las diferencias observadas dependen de la demanda cognitiva de dichas tareas.

Palabras clave: *evaluación, trabajo cooperativo, evaluación formativa, autorregulación, “zona de desarrollo próximo”.*

1. Introducción

Desde la Declaración de Bolonia en 1999, los distintos acuerdos encaminados a la creación de un Espacio Europeo de Educación Superior se concretan en el concepto de crédito ECTS (*European Credit Transfer System*) cuyo desarrollo implica, más allá de aspectos tales como el reconocimiento mutuo de las titulaciones, profundos cambios en la concepción del aprendizaje y el papel del profesorado. Uno de los retos esenciales consiste en situar la actividad del alumno como eje de un proceso de enseñanza-aprendizaje dirigido a la adquisición de un conjunto de competencias instrumentales, interpersonales y sistémicas adaptadas a los perfiles de cada titulación. Esta nueva orientación de la Educación Superior afecta de modo

sustantivo a los fines y estrategias de evaluación. En efecto, como ya se destacó en la introducción general, es un hecho claramente constatable, tanto desde la literatura didáctica como desde la propia experiencia docente, el poderoso impacto de la evaluación en las diversas facetas del aprendizaje (Crooks, 1988). De hecho, se puede afirmar que las estrategias y métodos de evaluación, de un modo más o menos explícito, determinan decisivamente los propios fines de la acción educativa.

En cuanto a las materias científicas y a la química en particular, si analizamos las competencias descritas en el documento *Tuning Educational Structures in Europe* (Mitchell y Whewell, 2003) existe un conjunto de problemas específicos que afectan al diseño del aprendizaje y su evaluación:

- La necesidad de lograr un equilibrio entre el “saber” y el “saber hacer”.
- La integración de los diversos planos epistemológicos que concurren en la química como disciplina científica. Por una parte, la realidad macroscópica, que puede abordarse de un modo descriptivo y ser objeto de estudio experimental; por otra, la interpretación de dicha realidad desde los diferentes modelos sobre la estructura de la materia y, por último, el nivel simbólico, relacionado con el lenguaje específico y el empleo de convencionalismos y representaciones gráficas (Gabel, 1999).
- La necesidad de contextualizar los contenidos del aprendizaje, tanto desde el punto de vista social y tecnológico como en relación al perfil profesional de la titulación.

Las exigencias que suscitan los actuales retos de la enseñanza universitaria y la consideración de las características específicas del aprendizaje de la química que se han descrito permiten, desde un análisis crítico de la situación actual, expresar algunas inquietudes que fundamentan los objetivos de este trabajo:

- La pertinencia y validez de las tareas de evaluación tradicionalmente utilizadas en la enseñanza universitaria es cuestionada por autores como Boussada y De Ketele, (2005). En efecto, su ámbito queda casi o totalmente restringido a categorías relacionadas, en términos de la taxonomía de Bloom, con el conocimiento y comprensión de hechos, conceptos y principios, así como en su aplicación a ejercicios numéricos, generalmente de un modo puramente algorítmico. Por este motivo, la incorporación efectiva de contenidos de carácter procedimental y actitudinal constituye un importante reto desde el punto de vista de la evaluación, tanto en sus estrategias globales como en sus métodos e instrumentos.
- Esta extensión de los contenidos del aprendizaje, consustancial al desarrollo de las competencias propias de cada titulación, exige abordar un importante

reto metodológico: la implicación del alumnado en la asignatura, transformando la clase en un espacio de indagación y reflexión, más allá de la mera transferencia de información. En definitiva, la incorporación de nuevos enfoques didácticos como el aprendizaje basado en problemas requiere extender las actividades de evaluación a contenidos de mayor complejidad y demanda cognitiva. Actividades relacionadas con procesos característicos de la metodología científica, desarrollo de actitudes, análisis de relaciones Ciencia-Técnica-Sociedad-Medio Ambiente (CTSA), etc., serían ejemplos en este sentido.

- La necesidad de dar respuesta a un alumnado notablemente heterogéneo, tanto en capacidad y en conocimientos previos, como en actitudes. Consecuentemente, es necesario conocer y asumir su situación inicial, dando prioridad al carácter de la evaluación como instrumento orientador del aprendizaje.
- Por último, y desde un punto de vista transversal, ya que afecta a todas las inquietudes expresadas anteriormente, cabe destacar la necesidad de impulsar en nuestras aulas el trabajo cooperativo, no sólo por razones metodológicas, sino por ser el fundamento de muchas de las competencias que nuestro alumnado debe alcanzar.
- Estas reflexiones han motivado la introducción de diferentes innovaciones en la asignatura *Química General*, del primer curso de Mecanización Agraria y Construcciones Rurales, de la Escuela Técnica Superior del Medio Rural y Enología (ETSMRE) de la Universidad Politécnica de Valencia, recogidas en Llorens, Llopis, y Edwards (2005). Otras experiencias similares en el ámbito de la Química Analítica han sido recientemente descritas (González-Casado, Valencia, Crices, Navas y Bosque, 2005; Del Olmo, García-Campaña, Gámiz-Gracia, Navalón y Blanc, 2005). El presente trabajo se refiere exclusivamente a la evaluación como instrumento de aprendizaje y a la introducción del trabajo cooperativo.

2. Objetivos

Partiendo de las inquietudes anteriormente descritas, podemos enunciar los siguientes objetivos específicos:

- El estudio de la influencia de la evaluación formativa en el rendimiento final de los estudiantes.
- El análisis, desde del concepto de “zona de desarrollo próximo” de Vygotsky, de la influencia del trabajo en pequeños grupos en la realización de actividades de evaluación.

- El desarrollo de instrumentos para la evaluación de los trabajos realizados en grupo, relacionados con la adquisición, procesamiento y presentación de la información.
- El análisis de la percepción por el alumnado acerca de las innovaciones introducidas y de los hábitos espontáneos de trabajo en grupo.

3. Modelo de aprendizaje en el que se sitúa esta propuesta

3.1. Consideraciones previas

Para fundamentar un modelo de evaluación que pretenda dar respuesta a los objetivos definidos es necesario partir de un marco teórico sobre el aprendizaje. Para establecerlo, puede ser útil considerar las condiciones que, según White (1999), favorecen un aprendizaje de calidad. Podrían sintetizarse así:

- La integración de los contenidos en una estructura coherente, construida a partir del establecimiento de relaciones significativas entre conceptos, que sea al mismo tiempo compatible con una adecuada diversificación de los métodos de enseñanza y evaluación.
- La consideración del papel central del trabajo del alumno y el desarrollo de su autonomía.
- La necesidad de promover el aprendizaje cooperativo, consecuente con la naturaleza social de la construcción del conocimiento.

3.2. Organización y estructura del contenido y su relación con la naturaleza y evolución de los conceptos científicos

La selección, delimitación y organización de los conceptos, así como el establecimiento de interrelaciones adecuadas entre ellos es uno de los elementos decisivos en la planificación docente. Según Galagovsky (2004), es necesario que el entramado conceptual gire en torno a ciertos “conceptos sostén” que den lugar al denominado “aprendizaje sustentable”, concepto que introduce como superación de la tradicional contraposición entre “aprendizaje memorístico” y “aprendizaje significativo” y como fruto de un análisis crítico de este último¹. En efecto, una de

¹ En la teoría del aprendizaje verbal significativo de Ausubel se considera como criterio el establecimiento de relaciones entre los nuevos conocimientos y la estructura cognitiva del que aprende. Desde este punto de vista, el que un aprendizaje sea significativo no implica necesariamente que sea correcto y contribuya a una evolución positiva. De ahí la introducción por la citada autora del concepto de aprendizaje sustentable como aquél que, siendo significativa, contribuye a una evolución positiva.

las deficiencias más características que afectan a la selección, organización y secuenciación de los contenidos es la ausencia de jerarquización conceptual y el carácter lineal y acumulativo con que se presenta el conocimiento científico (Gil, 1993).

3.3. Metodología basada en los procesos de resolución de problemas

Los modelos de cambio conceptual, que tanto impacto tuvieron hacia la década de los años 80, han ido evolucionando hacia posturas menos centradas en la identificación de las ideas erróneas y en la provocación del conflicto cognitivo como punto de partida para la construcción de conocimiento. Por el contrario, se pone cada vez un mayor énfasis en el establecimiento de contextos de aprendizaje caracterizados por la atención a la actividad metacognitiva (Parolo, Barbieri y Chrobak 2004), así como a la creación de situaciones problemáticas, generadoras de contextos comunicativos que proporcionen al alumno oportunidades de aplicar con éxito los conocimientos introducidos, consolidando su adquisición (Birch, 1986; Campanario y Moya, 1999). En la formación de futuros ingenieros, estas concepciones sobre el aprendizaje han derivado hacia la propuesta de modelos que ponen el acento, además, en el desarrollo del aprendizaje autónomo y de tareas de elevado nivel cognitivo que proporcionen una realimentación eficaz (Felder y Brent, 2004). En el terreno del aprendizaje científico, dichas tareas implican procesos de resolución de problemas, en su sentido más amplio, caracterizados por la presencia de rasgos característicos de la actividad científica: formulación y contrastación de hipótesis, propuesta de diseños experimentales, etc. Esta corriente podría enmarcarse dentro de lo que actualmente se denomina *Aprendizaje Basado en Problemas*.

3.4. Hacia un modelo de trabajo cooperativo

Desde el análisis tanto de las situaciones reales de aprendizaje como del tipo de competencias profesionales que son exigibles a un ingeniero, la relevancia del trabajo cooperativo nos lleva a dirigir nuestra atención al concepto de “zona de desarrollo próximo”, central en la psicología de Vygotsky, cuya actualidad puede considerarse claramente vigente en las actuales orientaciones socioconstructivistas del aprendizaje.

En efecto, tal como expone Paulson (1999), una estrategia metodológica basada en el planteamiento de problemas abiertos requiere avanzar hacia la generalización del trabajo cooperativo. Éste va mucho más allá de lo que habitualmente ha venido denominándose trabajo en equipo (Ovejero, 2005; Gómez Mújica y Acosta Rodríguez, 2005). A grandes rasgos, esta diferencia vendría dada por el hecho de que el verdadero trabajo cooperativo supone, más que una suma de aportaciones, un

cambio cualitativo que implica logros de mayor nivel cognitivo. Constatar este hecho y analizar qué factores contribuyen a dicho cambio es una línea de investigación relevante reflejada en trabajos como el de Punk y George (2003), en el que además de constatar la influencia positiva del trabajo cooperativo en el aprendizaje, se describe detalladamente un modelo de evaluación que considera la contribución de los miembros del grupo al trabajo colectivo mediante la evaluación entre iguales (*peer assessment*).

Otro objetivo esencial del trabajo cooperativo es la creación de situaciones comunicativas adecuadas. La necesidad de potenciar las habilidades argumentativas es propuesta por diversos autores, por ejemplo, Osborne, Erduran y Simon (2004) o Cordero, Colinvaux y Dumrauf (2002:427), quienes afirman que compartir procesos cognitivos a través del razonamiento y la argumentación, promotores de la construcción y reconstrucción de conocimientos científicos en grupos de alumnos, puede ser una respuesta a los desafíos de la práctica docente universitaria.

Dentro de la enseñanza de la química en los niveles iniciales universitarios existe una abundante literatura referida a la influencia del trabajo cooperativo en aspectos tales como la eficacia del cambio conceptual y la influencia del liderazgo (Basili y Sandford, 1991), el desarrollo de actitudes (Shibley y Zimmaro, 2002) o el desarrollo de la creatividad (Tang, Gallo, Jacobs y Kim-Eng, 1999). Existe además una creciente atención hacia el estudio de la dinámica interna de los grupos de trabajo. Por ejemplo, Haller, Gallagher, Weldon y Felder (2000) han analizado detalladamente, en estudiantes de ingeniería, la naturaleza de las interacciones verbales entre los distintos miembros de un grupo durante sesiones de aprendizaje cooperativo. A partir del análisis de los *roles* desempeñados por dichos miembros proponen pautas de intervención por parte del profesorado que pueden favorecer sensiblemente el desarrollo del trabajo cooperativo.

3.5. Estrategias y métodos de evaluación coherentes con los principios metodológicos expuestos

Los principios metodológicos expuestos exigen un modelo de evaluación que, partiendo de una comprensión profunda de la estructura del contenido, sea capaz de distinguir mejor lo esencial de lo accesorio, así como de identificar los conceptos y procedimientos que condicionan más claramente el resto de aprendizajes. Del mismo modo, debe abordar los procesos, no solamente los productos del aprendizaje, lo que lleva a acentuar su carácter diagnóstico y formativo.

En cuanto al papel del alumnado en la evaluación, la propuesta planteada en este trabajo pretende situarse en la perspectiva de la denominada “evaluación auténtica” (Astin, 2003). Este concepto de evaluación puede definirse, entre otros, por los siguientes rasgos:

- Implica a los alumnos en tareas que son percibidas por ellos mismos como significativas y son objeto, por tanto, de un esfuerzo de autoevaluación y autorregulación.
- Las actividades de evaluación son en sí mismas y se perciben como actividades de aprendizaje, planteándose dentro de secuencias y contextos donde se asegura una adecuada realimentación.
- Desde la propia actividad de evaluación debe hacerse partícipe al alumno de qué significa realizar correctamente la tarea que le proponemos, explicando las reglas establecidas para juzgar la calidad del trabajo.

Este concepto de evaluación debe enmarcarse en la influencia de las teorías cognitivas (Glaser, Raghavan y Baxter, 1992) y en la atención prestada a la actividad metacognitiva (Duschl y Gitomer, 1997; Baker, 1991). Una de las expresiones más representativas de este tipo de evaluación es el portafolio como instrumento de evaluación (Phelps y Laporte, 1997; Felder, 2005).

4. Materiales y métodos

4.1. La evaluación como práctica cotidiana

La propuesta que se presenta en este trabajo adopta como principio básico la pretensión de que la evaluación impregne el conjunto de todas las actividades de aprendizaje. Ello se concreta en la incorporación de prácticas educativas que permitan una realimentación lo más continua y eficaz posible. Así, por ejemplo, durante la clase se aplica frecuentemente la técnica *ConcepTest* (Kovac, 1999) en la cual cuestiones cuidadosamente seleccionadas permiten reorientar la clase. Del mismo modo, las actividades de autoevaluación son parte importante del material proporcionado a los alumnos. Desde una perspectiva más amplia, se exponen a continuación las actividades de evaluación desarrolladas y las innovaciones introducidas.

4.2. Innovación en las actividades de evaluación

4.2.1. Instrumentos de evaluación utilizados que han contribuido a la evaluación sumativa

En la tabla 1 se relaciona el conjunto de actividades de evaluación, el número de alumnos que han participado en cada una y la contribución porcentual a la calificación final.

PRUEBA	N	Impacto en la calificación final ⁽¹⁾
Evaluación de conceptos y procedimientos relacionados con la resolución de problemas		
Prueba inicial y su reelaboración	33	5 %
Test 1	24	5 %
Test 2	48	5 %
Prueba parcial 1	30	25 %
Prueba parcial 2 ⁽²⁾	34	20 %
Evaluación del trabajo experimental		
Test previo	32	5%
Examen de prácticas	27	10%
Cuaderno	24	5%
Trabajos monográficos en grupo (6 grupos de trabajo)		
Informe tema CTS	28	5%
Resumen artículo	28	5%
Revisión bibliográfica	28	5%
Presentación con diapositivas	28	5%
Cuestionario de valoración final		
Autoevaluación por el alumnado sobre su trabajo en grupo y valoración de aspectos organizativos	30	

(1) En el cálculo de la calificación final la corrección debida a los tests no realizados se aplica obteniendo la puntuación proporcional correspondiente a las pruebas realizadas.

(2) Consta de dos partes: una común con la materia nueva impartida (temas 4 y 5) y recuperación por los alumnos que no superaron la primera.

Tabla 1. Desglose de la evaluación sumativa

4.2.2. Estructura y contenido de las pruebas de evaluación de conceptos y su aplicación a problemas numéricos

- Evaluación inicial. Se lleva a cabo a través de una prueba de opción múltiple que posteriormente se reelabora por los alumnos como trabajo no presencial, comentando posteriormente en clase las alternativas erróneas de cada pregunta. Para optimizar la realimentación de esta prueba, dicha reelaboración también se corrige y califica. Los contenidos de la prueba se describen para cada uno de sus ítems en la exposición de sus resultados (Figura 2).

- Test 1. Consta de quince cuestiones de opción múltiple sobre el tema 1 (Estructura atómica y Sistema Periódico de los elementos).
- Test 2. Consta de quince cuestiones de opción múltiple acerca del tema 2 (Enlace químico. Relación entre la estructura y las propiedades de las sustancias).

Esta prueba se realiza en dos fases: individual y en grupos de 3-4 alumnos (las respuestas correctas se proporcionan inmediatamente después de ambas fases, bien en clase o a través de la microweb de la signatura). Ambas se califican por separado de manera que el rendimiento del trabajo colectivo puede modificar la calificación del individual. La variación en las calificaciones individuales como resultado de la resolución de la prueba en grupo viene dada por la siguiente expresión:

$$\Delta x = [x_g - \sum(x_i / N)]/N$$

Donde Δx es el incremento (positivo o negativo) aplicado a cada alumno, x_g es la calificación del ejercicio realizado en equipo, x_i las calificaciones individuales de los miembros del grupo y N su número.

- Prueba parcial 1. Comprende los dos primeros temas, así como los requisitos previos del curso (formulación y cálculos estequiométricos). Su estructura es la siguiente: 20 cuestiones abiertas de respuesta corta (50% de la prueba), 2 problemas (35%) y 20 cuestiones de formulación (15%).
- Prueba parcial 2. Su contenido se refiere a los temas 4 (Cambios energéticos en la reacción química. Los combustibles) y 5 (Equilibrio químico). Tiene lugar en la fecha de la convocatoria oficial y consta de 5 cuestiones estructuradas (60%) y 2 problemas (40%). El alumnado que alcanzó menos de un 50% o no se presentó al parcial 1 tuvo la oportunidad de realizar una prueba de estructura similar, pero referida a todo el programa de la asignatura.

4.2.3. Evaluación del trabajo experimental

El trabajo experimental se desarrolla mediante tres tipos de actividades:

- Una prueba rápida de 5 a 10 cuestiones de opción múltiple al comienzo de cada sesión de prácticas.
- La evaluación del cuaderno y la observación del trabajo de laboratorio (ver Apéndice A).
- Una prueba de cinco cuestiones estructuradas acerca de los trabajos experimentales en la que pueden utilizar su cuaderno de laboratorio.

4.2.4. Trabajos monográficos realizados en pequeños grupos

Los trabajos monográficos constituyen el modo de impartir el tema 3 del programa (Estudio específico de materiales de interés tecnológico: aleaciones y polímeros). Se realiza en 6 grupos de 3-4 alumnos². La propuesta de trabajo formulada agrupa cuatro tareas diferentes, basadas en la aplicación de una determinada técnica de trabajo:

- Realización de un informe, con su resumen, acerca de un tema de aplicación agrícola relacionado con los contenidos del tema.
- Análisis bibliográfico: elaboración de fichas bibliográficas sobre temas similares a los citados anteriormente.
- Comentario, resumen y contestación a un cuestionario sobre un artículo científico de divulgación, relacionado con los contenidos del tema y basado en las relaciones Ciencia-Técnica-Sociedad-Medio Ambiente (CTSA).
- Elaboración de una presentación en *power-point* sobre una de las actividades experimentales vinculadas al tema 3 (Determinación espectrofotométrica del acero, Corrosión del acero ó Síntesis y reconocimiento de polímeros).

El trabajo es propuesto al grupo durante una entrevista, estableciendo una reunión de seguimiento en tutoría. Todos estos trabajos son evaluados mediante una parrilla que se hace pública a principio de curso (ver Apéndice B).

4.2.5. Cuestionario para la valoración por el alumnado de las innovaciones introducidas

El cuestionario es aplicado al concluir el examen final y consta de dos grupos de cuestiones (ver Apéndice C):

- Valoración del trabajo en grupos:
 - Adecuación del volumen e interés del trabajo solicitado.
 - Criterios de calificación.
 - Organización y orientación. Medios.
 - Metodología desarrollada por el grupo.
- Trabajo experimental:
 - Interés, organización y orientación. Medios.
 - Utilidad del test inicial de cada sesión experimental.

² Fue necesario permitir la presentación de trabajos individuales a 5 alumnos que por razones laborales se presentaron directamente a una prueba final.

4.2.6. Evaluación y organización global del curso

En la figura 1 se describe de modo sintético y global la estrategia de evaluación aplicada y su relación con el desarrollo del curso.

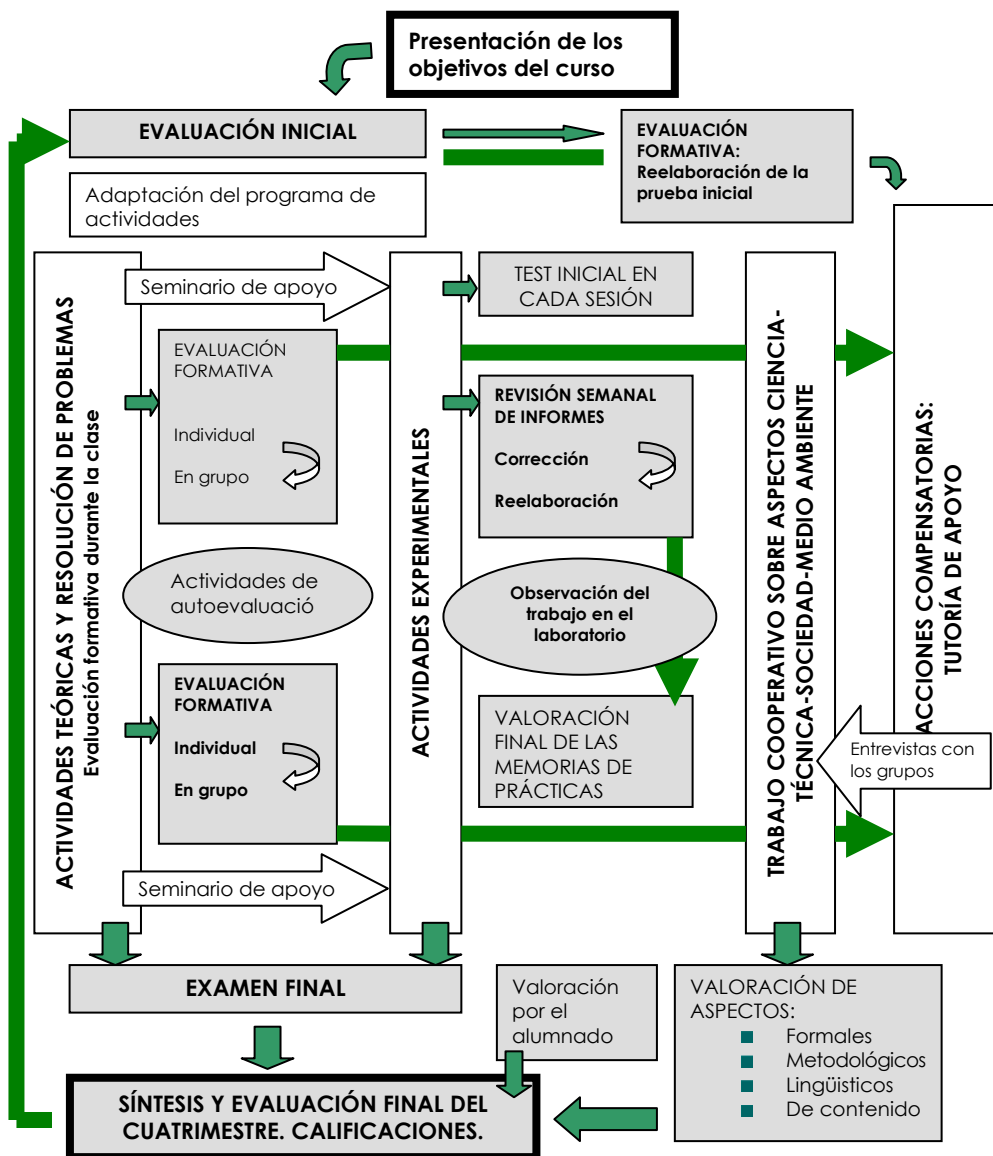


Figura 1. Estrategia de evaluación en la asignatura

4.3. Estudios realizados

4.3.1. *Análisis inicial. Influencia de la evaluación formativa*

En primer lugar, se pretende profundizar en el diagnóstico de la situación inicial y en cómo ésta es modificada por las tareas de aprendizaje propuestas. Para ello se estudia la evolución de algunos de los principales contenidos del curso a lo largo de las distintas pruebas realizadas.

Para estudiar la influencia de la evaluación inicial y formativa en la calificación global del curso y en la de sus diferentes aspectos (teoría y problemas, trabajo experimental y trabajos monográficos) se definen, a posteriori, dos subgrupos entre los alumnos a los que se otorga calificación final. Estos subgrupos se definen con el siguiente criterio:

- Alumnado que desarrolló evaluación formativa; se incluyen los que realizaron el Test 2 y la prueba parcial 1³.
- El resto de alumnos presentados al examen final.

En los análisis comparativos realizados la variable utilizada ha sido el índice de dificultad (ID) definido como el cociente, para un ítem determinado, entre el número de aciertos y el número total de respuestas. La significación de la diferencia entre los valores promedio se ha determinado con la prueba T de Student mediante la aplicación *Microsoft Excel*. Dicha prueba proporciona la probabilidad de que se cumpla la hipótesis nula consistente en admitir que los valores promedio comparados no son significativamente distintos, representando a la misma población.

4.3.2. *Influencia del trabajo cooperativo en la realización de tareas. Influencia del nivel cognitivo*

Se pretende estudiar la influencia en el aprendizaje de los procesos de autorregulación y de interacción entre iguales que tienen lugar en un contexto de trabajo cooperativo. Para ello se ha realizado un análisis comparativo entre el rendimiento individual y colectivo en la realización de tareas relacionadas con pruebas de opción múltiple. Se pretende contrastar la hipótesis, inspirada en el concepto de *zona de desarrollo próximo* de Vygotsky, de que la interacción entre los miembros del grupo es tanto más eficaz cuanto mayor es la complejidad cognitiva de la tarea.

³ La mayoría también realizó la prueba inicial y el test 1, pero se adoptó este criterio dado que algunos alumnos se incorporaron tarde al curso y no las realizaron; se integraron completamente con una asistencia regular.

De este modo, los ítems utilizados en este estudio han sido clasificados, tal como propone Jacobs (2000) en un estudio similar, en dos categorías:

Tipo 1: ítems que pueden relacionarse principalmente con las categorías A, D y E⁴ de la taxonomía propuesta por Klopfer (1975) que implican relaciones causales complejas y que no pueden resolverse por aplicación directa de actividades realizadas previamente en clase. Por ejemplo, las referidas a propiedades que dependen de la influencia simultánea de dos o más factores, como la siguiente pregunta.

¿Qué dos elementos forman un compuesto con mayor carácter iónico?

- a) Cs y F b) Na y F c) Cs y I d) Na y I

La resolución de este ítem implica relacionar dos ideas. Por un lado, el carácter iónico depende de la diferencia de electronegatividad entre los elementos y, por otro, la variación de la electronegatividad en la tabla periódica, que a su vez depende de la variación del número atómico (Z) a lo largo de un grupo y de un periodo.

Tipo 2: ítems relacionados con algunas subcategorías de A (conocimiento y comprensión) y que requieren procesos de menor nivel cognitivo⁵ tales como la selección de una definición, el reconocimiento de hechos específicos, establecer relaciones causales simples u otras más complejas pero que puedan resolverse de modo análogo a partir de actividades realizadas en clase, etc. Por ejemplo:

Entendemos por entalpía reticular...

- a) La energía liberada al formarse 1 mol de un sólido iónico cristalino a partir de los elementos en estado estándar.
b) La energía liberada al formarse 1 mol de un sólido cristalino a partir de los iones en estado gaseoso.
c) La energía necesaria para separar los elementos que forman 1 mol de una sustancia iónica.
d) La energía necesaria para fundir 1 mol de una sustancia iónica en estado sólido.

⁴ A: conocimiento y comprensión; D: interpretación de datos y formulación de generalizaciones; E: construcción prueba y revisión de un modelo teórico.

⁵ Esta menor demanda cognitiva no tiene por qué implicar necesariamente que el ítem sea más fácil y de hecho no se observa ninguna correspondencia en este sentido.

Es importante especificar que para la realización de los cálculos se ha adoptado como criterio restrictivo la selección de los ítems que cumplen las siguientes condiciones:

- Que puedan clasificarse de un modo razonablemente unívoco en una de las dos categorías.
- Que posean, en su aplicación individual o en grupo, un índice de discriminación superior a 0,2.
- Que no tengan índices de dificultad en la prueba individual muy extremos (mayores de 0,9 o menores de 0,1).
- Que no se haya constatado ningún problema de redacción o interpretación durante la aplicación de la prueba.

Este estudio se realiza en el test 2 y sus resultados se contrastan con una experiencia análoga realizada en un curso de 1º en la asignatura de *Fundamentos Químicos de la Ingeniería (Química Orgánica)*. En ambos se comparan (ver Tabla 2) los índices de dificultad (ID) de los ítems cuando son resueltos individualmente (ID_I) y en grupo (ID_G), aplicando la prueba t de Student para evaluar la significación de la diferencia entre las medias respectivas. Este análisis se realiza también por separado para los ítems del tipo 1 y del tipo 2. Al mismo tiempo, se considera también para cada ítem la diferencia entre los índices de dificultad cuando se aplica en grupo e individualmente ($ID_G - ID_I$). El significado de esta variable estaría relacionado con el avance cognitivo derivado de las interacciones entre los miembros del grupo. Los valores medios de estas diferencias para los ítems tipo 1 y tipo 2 se comparan también mediante la prueba T de Student con el fin de contrastar la hipótesis establecida.

4.3.3. Análisis comparativo de los distintos aspectos de los trabajos monográficos realizados en grupo

En cuanto a la evaluación de los trabajos monográficos, se realiza un análisis comparativo desde una doble perspectiva: por una parte, entre las cuatro tareas desarrolladas y, por otra, entre algunos aspectos transversales comunes a dichas tareas, concretamente:

- Valoración del contenido científico.
- Aspectos formales y de presentación.
- Adecuación a las normas y orientaciones propuestas.
- Uso del lenguaje, tanto desde el punto de vista semántico (precisión), ortográfico y sintáctico, como desde la cohesión de los textos elaborados.
- Empleo y citación de la bibliografía.

4.3.4. Valoración del curso

Por último, algunos de los aspectos más innovadores del diseño del curso fueron sometidos a su evaluación por el alumnado presentados al examen final. En este trabajo se presentan los principales resultados relacionados con el trabajo cooperativo, concretamente, la percepción por el propio alumnado del trabajo cooperativo y de la metodología seguida en su desarrollo.

5. Resultados

5.1. Prueba inicial y aplicación de los resultados como realimentación

Para la interpretación de los resultados de la prueba inicial, es necesario considerar que el nivel se sitúa en los conceptos básicos impartidos en Educación Secundaria. Dentro de un panorama general claramente deficiente (tan solo 2 ítems alcanzan un índice de dificultad del 50%), es especialmente destacable el pobre resultado alcanzado en aspectos tan básicos como los cálculos estequiométricos elementales o sobre los fundamentos de la estructura atómica (ver Figura 2).

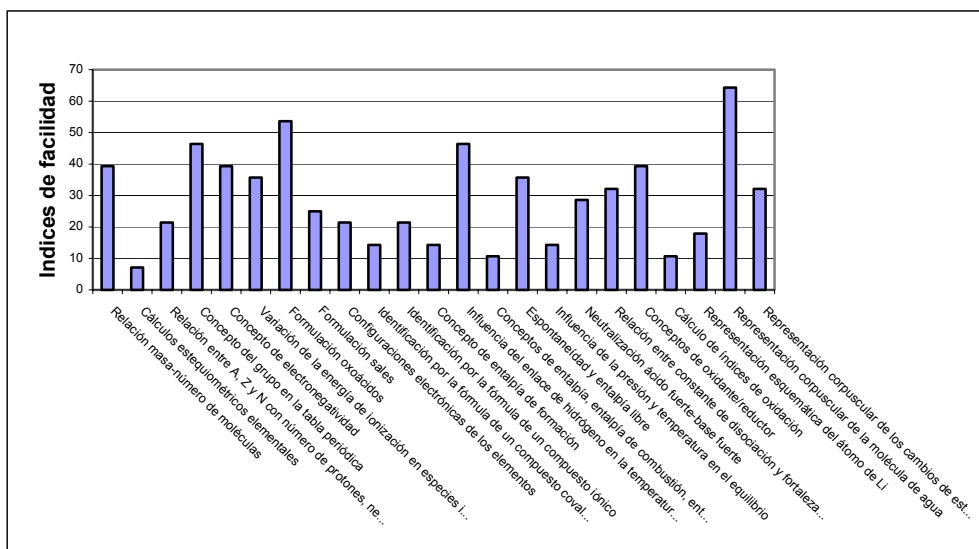


Figura 2. Puntuación por ítems de la prueba inicial

A la hora de valorar la efectividad de la realimentación se puede comparar, para los núcleos conceptuales fundamentales, el rendimiento alcanzado por aquellos alumnos que han realizado la prueba inicial, su reelaboración y los dos primeros tests (contenido relacionado con estructura y propiedades de la materia).

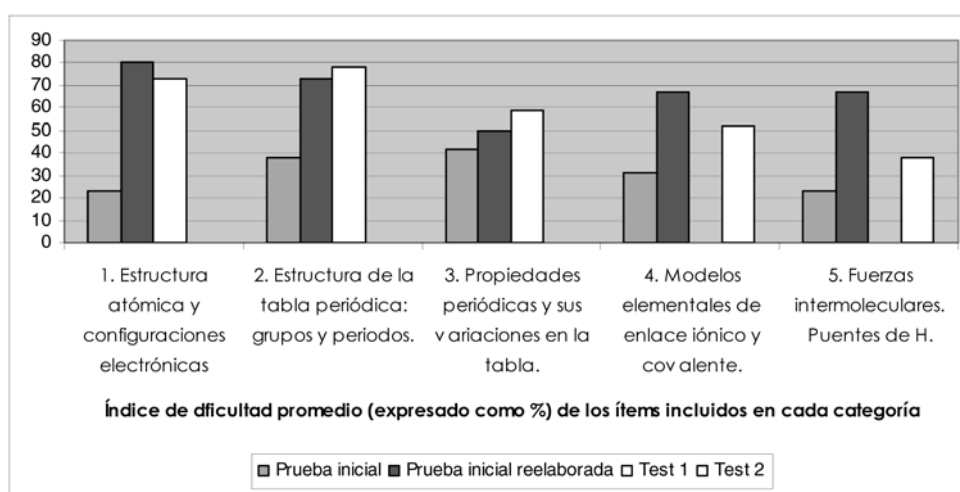


Figura 3. Evolución del aprendizaje de contenidos considerados como requisitos previos en el curso

En los resultados que aporta la figura 3 se observa un esperado y notorio incremento del nivel de aprendizaje al reelaborar la prueba (no es tan obvio si se considera que el criterio de evaluación es la argumentación adecuada, no la simple elección de la alternativa correcta). Lo más llamativo es observar cómo en los tests 1 y 2 se mantiene en mayor o menor medida dicho incremento, mejorándose incluso en algún caso, como, por ejemplo, en la estructura de la tabla periódica de los elementos.

5.2. Evaluación formativa y resultados globales

En cuanto a la eficacia de la evaluación formativa, se comparan los resultados alcanzados en cada uno de los aspectos evaluados en el curso, considerando los dos subgrupos definidos (alumnado que desarrolló o no evaluación formativa). El primer subgrupo alcanza un 72,7% de aprobados frente a un 9,1% en el segundo. Analizando la significación de la diferencia entre las calificaciones globales medias según la prueba t de Student, se observa que ambas son diferentes con una probabilidad prácticamente del 100%, tal y como muestra la figura 4.

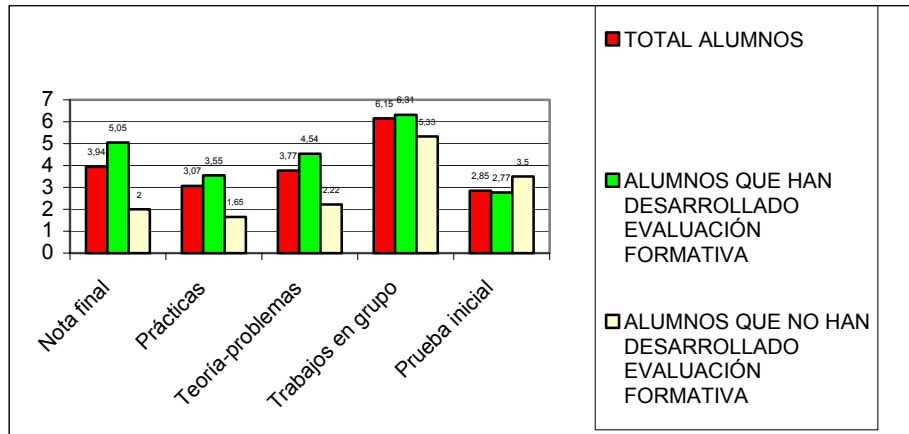


Figura 4. Desarrollo de evaluación formativa y rendimiento en los diferentes contenidos del curso

5.3. Comparación entre el rendimiento del trabajo individual y de grupo en tareas de evaluación. Influencia del nivel cognitivo de los ítems

En la comparación entre el rendimiento individual y de grupo se observa que la interacción entre los alumnos incide favorablemente y de un modo significativo en la resolución de la prueba. Tras aplicar la prueba t de Student a los resultados del trabajo individual y de grupo se obtiene una probabilidad de rechazar la hipótesis nula del orden del 99% (ver Tabla 2). Desde un punto de vista cualitativo se observa la creación de un clima de trabajo muy satisfactorio. Esta comparación es mucho más relevante cuando se considera la demanda cognitiva de los ítems, como reflejan los resultados mostrados en la figura 5.

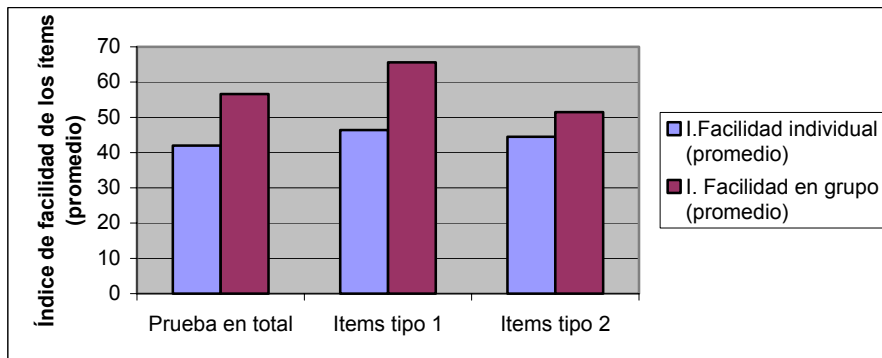


Figura 5. Diferencias entre la prueba individual y en grupo para los tests del tipo 1 (mayor nivel cognitivo) y los del tipo 2 (menor nivel cognitivo)

En los resultados de la tabla 2 se advierte, en primer lugar, que tanto para cada prueba por separado como considerando conjuntamente todos los ítems, las diferencias entre los valores medios de los índices de dificultad para la prueba individual y en grupo son significativas con probabilidades en torno al 1%, 18% y 1% de aceptar la hipótesis nula⁶, respectivamente.

N = 24 ítems	Valores promedio de los índices de dificultad ID_i (pruebas individuales) y ID_G (pruebas en grupo) y probabilidad (P) de la hipótesis nula para la prueba t de Student.								
	Prueba completa			Ítems Tipo 1 (N = 11)			Ítems Tipo 2 (N = 13)		
	P_i	P_G	$P_G - P_i$	P_i	P_G	$P_G - P_i$	P_i	P_G	$P_G - P_i$
Q.General	0,442	0,628	0,186	0,445	0,695	0,250	0,435	0,570	0,135
P	0,010			0,003			0,204		
P($ID_G - ID_i$)				0,039					
Q. Orgánica	0,550	0,727	0,177	0,490	0,734	0,244	0,600	0,720	0,120
P	0,182			0,308			0,473		
P($ID_G - ID_i$)				0,06					
Análisis conjunto	0,491	0,674	0,183	0,465	0,713	0,248	0,510	0,640	0,130
P	0,010			0,025			0,184		
P($ID_G - ID_i$)				0,003					

Tabla 2. Rendimiento individual y de grupo

Por otra parte, si se comparan directamente los índices de dificultad medios en la prueba individual y de grupo, respectivamente, para cada grupo de ítems, se observa que la probabilidad de aceptar la hipótesis nula (las medias no son significativamente diferentes) es mucho menor en los ítems del tipo 1 (0,030-0,204; 0,308-0,473; 0,025-0,184).

Los valores de la probabilidad asociada a la prueba t de Student muestran también que la diferencias entre los índices de dificultad promedio de los ítems resueltos individualmente y en grupo son, con una elevada probabilidad, significativamente más elevadas en los ítems tipo 1 (0,250; 0,244; 0,248) que en los ítems del tipo 2 (0,135; 0,120; 0,130).

⁶ Aceptar la hipótesis nula consiste precisamente en admitir que los valores medios no son significativamente distintos, representando a la misma población. Así, por ejemplo, un valor de $P = 0,10$ significaría que con una probabilidad del 90% podemos rechazar la hipótesis nula y admitir que las dos medias corresponden a poblaciones diferentes.

5.4. Trabajos en equipo

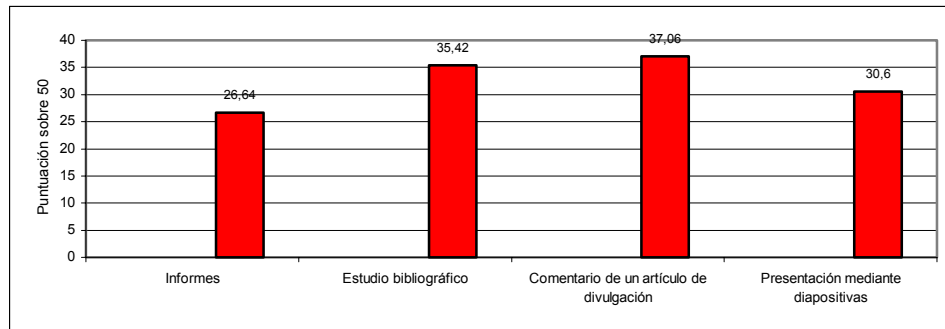


Figura 6. Puntuación media de los diferentes trabajos realizados en grupo

Tanto desde la percepción del alumnado como desde sus resultados académicos, se observa que los trabajos monográficos en equipo son las tareas de mayor aceptación, en particular el trabajo con artículos y textos escritos en general (ver Figura 6). Si se agrupan los aspectos comunes de cada tipo de tarea (ver Figura 7), en los resultados se observa que la valoración del contenido científico es menor que la de los aspectos formales y lingüísticos (la significación de este último aspecto es dudosa por la abundancia de transcripciones literales de las fuentes bibliográficas consultadas). Cabe destacar el deficiente empleo de la bibliografía, tanto en su grado de utilización como en su calidad y pertinencia: ausencia de revistas, páginas web poco específicas y seleccionadas, etc.

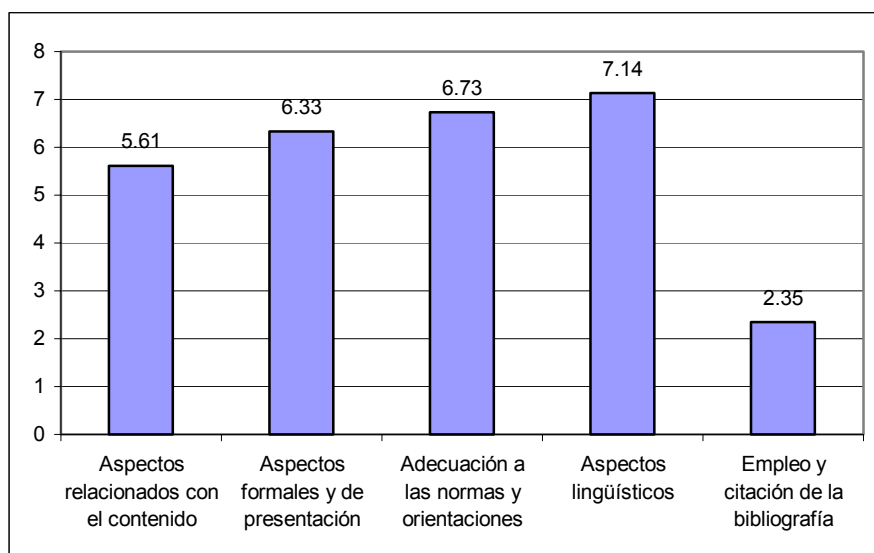


Figura 7. Evaluación de algunos aspectos transversales de los distintos tipos de tareas en los trabajos monográficos en grupo

5.5. Valoración por el alumnado de algunas de las innovaciones introducidas

5.5.1. Sobre el trabajo en grupo

Se observa que la realización de trabajos en grupo es bien valorada en general (ver Figura 8), aunque a la luz de los comentarios abiertos cabe matizar que hay un núcleo significativo que más que apreciar el trabajo en grupo por su posible carácter cooperativo, lo valora por el propio contenido de los trabajos y la metodología empleada.

Un ejemplo significativo sería: *“He aprendido a buscar bibliografía. El trabajo en grupo me parece que para aprender a buscar datos y asimilarlos sí que sirve pero siempre hay problemas dentro del grupo. Un trabajo individual sería mejor. El sistema de evaluación me parece correcto”*.

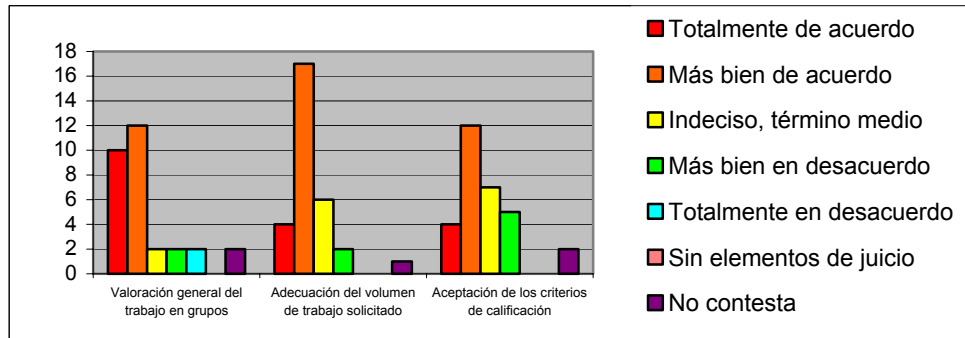


Figura 8. Sobre el trabajo en grupos (I). Valoración por el alumnado de algunas de las innovaciones introducidas

En cuanto a la metodología de trabajo en grupo (ver Figura 9) existe una elevada dispersión en la percepción de los alumnos. No se observa un modelo espontáneo definido de trabajo. Este hecho se aprecia con claridad si se comparan las contestaciones de los miembros de un mismo grupo. En principio, solamente hay dos grupos (de los seis que se constituyeron), donde se observa coherencia en su apreciación de cuál fue la metodología de trabajo seguida. Eran precisamente los grupos formados por los alumnos más implicados en la asignatura y que obtuvieron un mayor rendimiento.

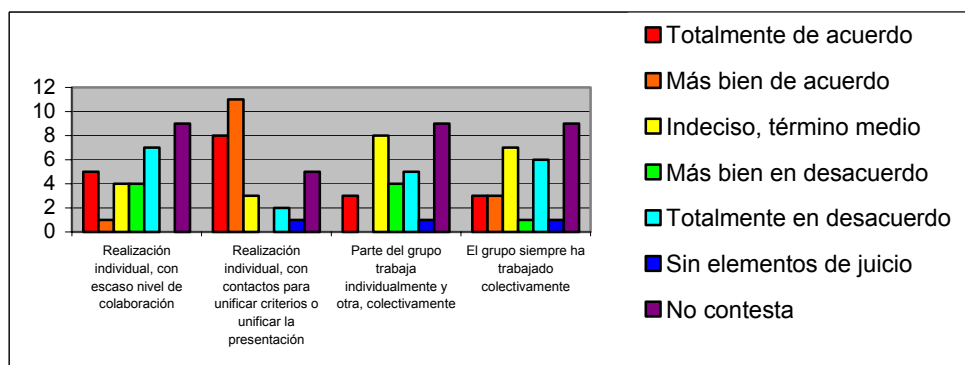


Figura 9. Sobre el trabajo en grupo (II). Metodología desarrollada por el grupo

La organización y orientación del trabajo (ver Figura 10) es bien valorada; por ejemplo: *“Ha mejorado mi aprendizaje en cuanto a buscar información fuera del material de clase, ya sea por internet o en bibliotecas, sobre todo en el tema del trabajo de bibliografía”*. No obstante, cabe destacar la elevada proporción de indecisos a la hora de pronunciarse sobre las orientaciones y medios disponibles.

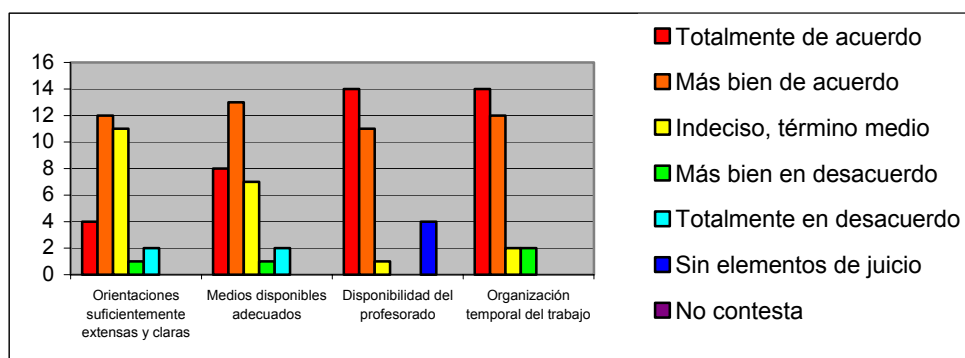


Figura 9. Sobre el trabajo en grupo (III): organización y orientación

6. Conclusiones

I. Los resultados demuestran que la implicación en las tareas cotidianas de la clase es el factor decisivo en el rendimiento académico, más que el nivel de conocimientos previos. Son principalmente los alumnos recién ingresados y que asisten habitualmente a clase aquellos que superan la asignatura mayoritariamente en la convocatoria ordinaria. Los suspensos corresponden principalmente al alumnado, muchas veces repetidor, que asiste ocasionalmente y no suele implicarse en las actividades de evaluación formativa.

Los resultados obtenidos, por modesto que sea su alcance, cuestionan la creencia, muchas veces aceptada acriticamente, de que el fracaso académico en los primeros cursos de la universidad se debe exclusivamente a carencias heredadas de la Educación Secundaria. En cualquier caso, proponer soluciones a este problema exige, en primer lugar, delimitar claramente su naturaleza y analizar los distintos factores que subyacen en él y, a partir de ahí, proponer iniciativas que favorezcan una integración adecuada del alumnado que accede a la universidad, como son la orientación adecuada en la Educación Secundaria, criterios de selección razonables, iniciativas para la adaptación de los niveles académicos iniciales, etc., y, por supuesto, unos criterios de permanencia que estimulen el compromiso personal de los estudiantes con su trabajo.

II. En cuanto al establecimiento del trabajo cooperativo, es notoria en muchos grupos la divergencia entre las percepciones de sus componentes acerca de su propio modo de trabajar. La inestabilidad en el funcionamiento de muchos grupos por las incorporaciones tardías y por los abandonos durante el curso es un factor que debe considerarse en la interpretación de estos resultados. Por otra parte, es evidente la necesidad de una acción tutorial específica para impulsar este tipo de trabajo.

III. El sistema de evaluación de la asignatura se percibe de manera prácticamente general como positivo, siendo los principales argumentos la valoración del trabajo continuado durante el curso y la diversificación de las actividades, que son apreciadas como oportunidades para realizar tareas poco habituales. Un ejemplo de este punto de vista es la siguiente afirmación formulada en el cuestionario final: *“Me ha parecido adecuado el método de evaluación, ya que el tener exámenes parciales de cada tema al estudiar, entiendes mejor el tema siguiente. Al hacer el estudio progresivamente se facilita la comprensión y el estudio, con lo que considero que se mejora el rendimiento”*. Tan solo aparece alguna crítica puntual dirigida a la excesiva complejidad que supone el cálculo de las calificaciones.

Por otra parte, es evidente la necesidad de profundizar en aspectos de la evaluación tales como la autoevaluación y la coevaluación o evaluación entre iguales como instrumentos para la mejora de las habilidades metacognitivas y del trabajo cooperativo.

IV. Otra conclusión relativa a la evaluación en general, y que deriva más de la reflexión sobre la práctica que de los datos obtenidos, es que contenidos básicos en la formación química de los alumnos acaban siendo poco determinantes en su calificación. Puede darse el caso, por ejemplo, de que un alumno con carencias importantes en formulación o cálculos estequiométricos pueda aprobar el curso. Posiblemente, los modelos de calificación de tipo aditivo basados fundamentalmente en la incorporación de más aspectos del aprendizaje a la nota final, aún aportando objetividad, siguen siendo insuficientes. Por el contrario, la introducción de un modelo más jerarquizado que incorpore la evaluación por criterios, además de proporcionar calificaciones más justas ayudaría al alumnado a identificar aquellos contenidos de carácter más fundamental para su formación. Evidentemente, el establecimiento de estos criterios exige un elevado nivel de coordinación en los equipos docentes.

V. Las actividades que implican procesos de autorregulación (reelaboración de la prueba inicial y de los exámenes en grupo) muestran una clara potencialidad, no tan sólo porque inciden en la mejora del aprendizaje, sino por el clima de trabajo que generan. En cuanto a la influencia del trabajo cooperativo en actividades de evaluación de cierto nivel cognitivo, se observa una clara mejora respecto de su

aplicación individual. Esta conclusión parece, en principio, obvia. No lo es tanto si observamos que la interacción entre iguales resulta ser más eficaz en las cuestiones de mayor requerimiento cognitivo. De hecho, en algunos casos se observa cómo el grupo escoge una opción que individualmente no había sido propuesta por ninguno de sus miembros. Estos resultados confirman en buena medida la relevancia del trabajo cooperativo y la posibilidad de centrar su campo de aplicación a las actividades cognitivamente más complejas. Tal y como ya se estableció en la fundamentación teórica de este trabajo, el concepto de *zona de desarrollo próximo* en la teoría de Vygotsky es crucial a la hora de diseñar actividades en las que el trabajo cooperativo muestre toda su potencialidad y eficacia. Caracterizar y delimitar dicha *zona de desarrollo próximo* es, por tanto, un importante objetivo en el diseño de procesos de aprendizaje. En este sentido, podría afirmarse que un componente importante de la misma en el contexto de un curso universitario de química general sería la introducción y aplicación de relaciones multivariantes entre conceptos.

Esta última conclusión puede abrir interesantes perspectivas para futuras investigaciones, ya que caracterizar en contextos educativos suficientemente representativos el tipo de actividades que definen la zona de desarrollo próximo en cada dominio de contenido puede ser una valiosa contribución a un diseño más eficiente de procesos de aprendizaje cooperativo.

Bibliografía

- Astin, A. W., Banta, T. W., Cross, K. P., El-Khawas, E., Ewell, P. T., Hutchings, P., Marchese, T. J., McKlenney, K. M., Mentkowski, M., Miller, M. A., Moran, E. T. y Wright, B. D. (2003). 9 principles of good practice for assessing student learning. American Association for Higher Education.
<http://www.aahe.org/aaheSearch.cfm> [ref. 16 abril 2005].
- Basili, J.C., Sandford, P.J., (1991). Conceptual change strategies and cooperative group work in chemistry. *Journal of Research in Science Teaching*, 28: 293-304.
- Birch, W. (1986). Towards a model for problem-based learning. *Studies in Higher Education*, 11: 73:82.
- Boussada, H. y De Ketele, J. M. (2005). *L'évaluation comme indicateur de la qualité de la formation universitaire: étude empirique dans le cadre du mastère de didactiques spécialisés*. 22^e Congrès de l'Association Internationale de Pédagogie Universitaire (AIPU), Ginebra, septiembre 2005.
- Campanario, J. M. y Moya, A. (1999). ¿Cómo enseñar ciencias? Principales tendencias y propuestas. *Enseñanza de las Ciencias*, 17(2): 179-192.

- Cordero, S., Colinvaux, D. y Dumrauf, A.G. (2002). ¿Y si trabajan en grupo...? Interacciones entre alumnos, procesos sociales y cognitivos en clases universitarias de Física. *Enseñanza de las Ciencias*, 20(3):427-442.
- Crooks, T.J., (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4): 438-481.
- Del Olmo, M., García-Campaña, A., Gámiz-Gracia, L., Navalón, A. y Blanc, R. (2005). *La Química Analítica en el Espacio Europeo de Enseñanza Superior. Una experiencia piloto en los estudios de ingeniería química*. VII Congreso Internacional sobre Didáctica de las Ciencias, Granada, septiembre de 2005.
- Duschl, R.A. y Gitomer, D.H. (1991). Epistemological perspectives on conceptual change: Implications for educational practice. *Journal of Research in Science Teaching*. 28(9): 839-858.
- Felder, R. (2005). *Active and cooperative learning*.
<http://www.ncsu.edu/felder-public/RMF.html> [ref. 14 abril 2005].
- Felder, R. M. y Brent, R., (2004). The intellectual development of science and engineering [students](#). 2. Teaching to promote growth. *Journal of Engineering Education*, 93(4): 279-291.
- Gabel, D. (1999). Teaching and learning chemistry education research: A look to the future. *Journal of Chemical Education*, 76 (4): 548-555.
- Galagovsky, L. R. (2004). Del aprendizaje significativo al aprendizaje sustentable (partes 1 y 2). *Enseñanza de las Ciencias*. 22, (1): 229-241; 22 (2): 349-364.
- Gil, D., (1993). Contribución de la historia y la filosofía de las ciencias al desarrollo de un modelo de enseñanza/aprendizaje como investigación. *Enseñanza de las Ciencias*, 11(2): 197-212.
- Glaser, R., Raghavan, K. y Baxter, G.P. (1992). *Cognitive theory as the basis for design of innovative assessment design characteristics of science assessments*. National Center for Research on Evaluation, Standards and Student Testing. University of California, Los Angeles.
<http://www.cse.ucla.edu/CRESST/Reports/TECH452.pdf> [ref. 18 marzo 2005].
- Gómez Mujica, A. y Acosta Rodríguez, H. (2003). Acerca del trabajo en grupos o equipos. *ACIMED (online)*; nov-dic. 2003, 11(6). [ref.12 junio 2005].
<http://scielo.sld.cu/scielo.php?pid=s1024-94352003000600011&scrip=sciabstract&tlng=es> [ref. 24 mayo 2005].

- González-Casado, A., Valencia, M., Crices, C., Navas, N. y Bosque, J. M. (2005). *Adaptación de la asignatura Química Analítica Instrumental al Espacio Europeo de Educación Superior (EEES)*. VII Congreso Internacional sobre Didáctica de las Ciencias. Granada (España), septiembre de 2005.
- Haller, C. R., Gallagher, V. J., Weldon, T. L. y Felder, R. M. (2000). Dynamics of peer education in cooperative learning workgroups. *Journal of Engineering Education*, 89 (3): 285-293.
- Jacobs, D.C. (2000). *Cooperative learning in General Chemistry*. Department of Chemistry and Biochemistry. Notre Dame, IN: University of Notre Dame.
- Klopfer, L. E. (1975). Evaluación en el aprendizaje de las ciencias. En Bloom, B., Hastings, T. y Madaus, G. (eds.). *Evaluación del aprendizaje*. Buenos Aires: Ed. Troquel.
- Kovak, J. (1999). Student active learning methods in General Chemistry. *Journal of Chemical Education*, 76(1): 120-124.
- Llorens, J.A., Llopis, R. y Edwards, M. E., (2005). *La adaptación al Espacio Europeo de Educación Superior: Organización de los contenidos y estrategias de evaluación en un curso de química general*. VII Congreso Internacional sobre investigación en didáctica de las ciencias, Granada, septiembre de 2005.
- Mitchell, T. y Whewell, R. (2003). Chemistry Subject Area Group: The Chemistry “Eurobachelor”. En *Tuning Educational Structures in Europe*. Comisión Europea de Educación y Cultura, 111-123.
- Munk, V. y George, A. V. (2003). Small group learning in first year chemistry: Does it work?. *Australian Journal of Education Chemistry*. 62: 9-12.
- Osborne, J., Erduran, S. y Simon, S., (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41: 994-1020.
- Ovejero, B. A. (2005). *Aprendizaje cooperativo*. Facultad de Psicología. Universidad de Oviedo.
www.psi.uji.es/asignatura/obtener.php?letra=R&codigo=26&fichero=1065533587R26-Resultado-Suplementario
[ref. 26 mayo 2005].
- Parolo, M. E., Barbieri, L.M. y Chrobak, R. (2004). La metacognición y el mejoramiento de la enseñanza de química universitaria. *Enseñanza de las Ciencias*, 22 (1): 79-92.

- Paulson, D. R., (1999). Active learning and cooperative learning in the Organic Chemistry lecture class. *Journal of Chemical Education*, 76(8): 1136-1140.
- Phelps, A. J., LaPorte, M. M. y Mahood, A. (1997). Portfolio assessment in high school Chemistry: One teacher's guidelines. *Journal of Chemical Education* 74(5): 528.
- Shibley, I. A. y Zimmaro, D. M. (2002). The Influence of collaborative learning on student attitudes and performance in an introductory chemistry laboratory. *Journal of Chemical Education*, 79: 745.
- Tang, G., Gallo, P. B., Jacobs, G. M. y Kim-Eng, C. (1999). Using cooperative learning to integrate thinking and information technology in a context-based writing lesson. *The Internet TESL Journal*. V(8)
<http://iteslj.org/Techniques/Tan-Cooperative.html>
- White, R. T. (1999). Condiciones para un aprendizaje de calidad en la enseñanza de las ciencias. *Enseñanza de las Ciencias*. 17 (1): 3-16.

Apéndice A. Documentos para la evaluación del trabajo experimental

Criterios de corrección del cuaderno de laboratorio

ALUMNOS	Realización de las actividades	Tablas y gráficas. Aplicación de normas	Presentación: orden, legibilidad y limpieza	Utilización de aplicaciones informáticas	Existen anotaciones sobre el desarrollo de la práctica, comentarios	Aparecen dibujos y esquemas de montajes	TOTAL/10 (5 PUNTOS)
	0-15	0-5	0-5	0-10	0-10	0-5	
::							

Documento para la evaluación del trabajo en el laboratorio

Fecha.....Práctica.....Grupo.....Subgrupo.....

1 A	Registro de incidencias	1 C	Registro de incidencias
1 B	Notas:	1D	Notas:

2 A	Registro de incidencias	2 C	Registro de incidencias
2 B	Notas:	2 D	Notas:

La evaluación compartida: investigación multidisciplinar

Las claves 1 A, 2 A, etc., corresponden a la localización de cada una de las parejas de prácticas. (El documento puede ampliarse según el número de alumnos). Esta denominación se aplica también a las bandejas de material asignadas a cada grupo. En este documento aparecen los nombres de los alumnos correspondientes, los cuales, salvo ausencias, forman grupos estables a lo largo del curso. Las incidencias que son registradas habitualmente en clase son las expuestas a continuación. Cualquiera otra es anotada aparte.

1. No trae bata o cuaderno.
2. Realiza preguntas que denotan comprensión de la práctica e interés por aspectos relacionados.
3. Realiza preguntas que muestran una desconexión total con el desarrollo de la clase.
4. Utiliza correctamente el material y los equipos de laboratorio.
5. Maneja de manera incorrecta equipos y material, habiéndose explicado previamente.
6. Se observan deficiencias en orden y limpieza en su puesto de trabajo.
7. Realiza las actividades previas indicadas en el cuaderno.
8. Lleva al día el cuaderno de prácticas.

Apéndice C. Cuestionario para la evaluación por el alumnado de distintos aspectos del curso

Nombre y apellidos

El objetivo de este cuestionario es valorar algunos aspectos del enfoque del curso para poder mejorar su organización y contenido. Sus resultados NO TIENEN NINGUNA REPERCUSIÓN EN LA CALIFICACIÓN.

	Totalmente de acuerdo	Más bien de acuerdo	Indeciso, término medio	Más bien en desacuerdo	Totalmente en desacuerdo	Sin elemento de juicio
Marcar (X) en la casilla correspondiente a la opción elegida						
Sobre el trabajo en grupos						
En general, la realización de trabajos en grupo la considero positiva en una asignatura de las características de Química General.						
El volumen de trabajo solicitado es adecuado.						
Los criterios de calificación propuestos me parecen justos y adecuados a los objetivos de la asignatura.						
Los temas propuestos han sido interesantes y adecuados.						
a) Informes.						
b) Revisión bibliográfica.						
c) Comentario de un artículo.						
d) Presentación o poster.						
Metodología del trabajo: ¿Con qué modelo de trabajo coincide más claramente el trabajo de tu grupo?						
El trabajo ha sido realizado individualmente, con muy poco nivel de colaboración.						
Los trabajos se dividen y se realizan individualmente pero al menos, al principio y al final, hay una toma de decisiones conjunta para unificar criterios o dar cierta unidad a la presentación final.						
Parte del grupo ha trabajado colectivamente (en colaboración) y otros miembros de manera individual.						
El grupo siempre ha trabajado colectivamente.						

Marcar (X) en la casilla correspondiente a la opción elegida	Totalmente de acuerdo	Mts bien de acuerdo	Indeciso, término medio	Mts bien en desacuerdo	Totalmente en desacuerdo	Sin elementos de juicio
Sobre la organización y orientaciones. Material y fuentes de consulta disponibles. Medios.						
Las orientaciones proporcionadas han sido suficientemente extensas y claras.						
Los medios disponibles han sido los adecuados.						
La disponibilidad del profesorado a través de las tutorías ha permitido recibir las aclaraciones con su orientaciones necesarias.						
La organización temporal: fechas de entrega de orientaciones, entrega de los trabajos, etc., ha sido la adecuada.						
Valoración del trabajo experimental						
El cuaderno proporcionado ha sido adecuado y útil.						
El test previo de cada práctica ha sido eficaz ya que nos ha obligado a emplear un poco de tiempo en la preparación de cada práctica.						
Los medios materiales son adecuados.						
El contenido de las prácticas me ha parecido interesante y útil.						
En general, sabía qué iba a hacer y por qué. Tenía claro el objetivo de cada práctica.						

Valoración general. Intenta resumir en unas pocas líneas QUÉ HAS APRENDIDO EN ESTA ASIGNATURA, no sólo en cuanto a contenidos sino en cuanto a métodos y técnicas de trabajo. Indica también si el SISTEMA DE EVALUACIÓN DE LA ASIGNATURA te ha ayudado a mejorar tu aprendizaje y por qué.



Evaluación continua y formativa en un curso universitario de química

*Miguel Anglés
Mónica Edwards
Rafael Llopis
Juan Llorens*

En este trabajo se presentan algunos resultados de una investigación sobre una propuesta de evaluación continua desarrollada durante tres años con estudiantes de primer curso de Química para Ingeniería. Dicha propuesta, de orientación constructivista y fundamentada en la teoría del procesamiento de la información y el cambio conceptual, se ha centrado en el trabajo cooperativo y en el diseño de actividades de evaluación integradas en todo el proceso de enseñanza-aprendizaje. Se detalla el alcance y efectividad de la propuesta así como la aplicación de un software específico (París), utilizado con el objetivo de optimizar el desarrollo y la autoevaluación. Los resultados obtenidos en este estudio muestran una notable mejora en el grado de participación, en el aprendizaje y, como consecuencia, en el rendimiento académico de los estudiantes.

Palabras clave: *Evaluación, rendimiento académico, EEES, trabajo cooperativo, tecnologías de la información y comunicación (TIC).*

1. Introducción

Los bajos niveles en el rendimiento académico de los estudiantes configuran una problemática recurrente en España y en otros países europeos, además de constituir una de las mayores preocupaciones desde el punto de vista de la pretendida armonización de los sistemas de educación superior europeo (Dearing, 1997; Attali, 1998; Yorke, 1999; Smith, 1999; Comisión de las Comunidades Europeas, 1995, 2003).

En la Universidad Politécnica de Valencia, así como en una gran mayoría de instituciones de enseñanza científica y técnica, los alumnos emplean en acabar su carrera un número de años muy superior al teórico establecido, con elevadas tasas de abandono (Bricall, 2000; Hernández Armenteros, 2002; Hernández, Bará, Córdoba, De Luis y Martín, 2002; Consejo de Coordinación Universitaria, 2003). Los índices de abandono y repetición son particularmente elevados en los primeros cursos; con frecuencia los alumnos no asisten a clase y muestran un escaso compromiso con su aprendizaje, falta de interés y motivación (Pozo, 2001; Marchesi, 2003). Esta situación suele atribuirse a deficiencias en el proceso de selección previo al ingreso en la Universidad, así como graves carencias en la formación precedente. No obstante, es pertinente señalar que se trata de una problemática compleja en la que incide una multiplicidad de factores, todavía no suficientemente investigada y, no en pocos casos, ni siquiera reconocida en su verdadera magnitud (Santos Guerra, 1999; Yorke, 1999; Campanario, 1998). Algunas revisiones, como las realizadas por Rodríguez Espinar (2003), muestran la existencia de numerosos estudios que describen los resultados académicos al final de un determinado período y del seguimiento. Pese a ello, aún no se ha producido un planteamiento comprensivo y profundo que permita explicar las dificultades de la transición de la educación secundaria a la universidad.

Por otra parte, el proceso de construcción del Espacio Europeo de Educación Superior (EEES) en que estamos inmersos representa una excelente oportunidad para tratar no sólo de resolver los endémicos problemas de rendimiento sino para promover un auténtico cambio de cultura académica. Un cambio que implica, indudablemente, tratar de poner fin a los elevados niveles de absentismo y abandono, a una excesiva centralización curricular en contenidos conceptuales, a la existencia de temarios interminables, al predominio de la clase expositiva, al papel pasivo del alumno, la evaluación reducida a un único examen final, entre otros (Fabra y Doménech, 2001; Campanario, 2002). Como sostienen los miembros de la *Commission on Higher Education* (1997: 94), “el rendimiento académico de los estudiantes es un componente clave para determinar si una institución está alcanzando sus objetivos educativos”. El proceso de Bolonia y los objetivos planteados brindan la ocasión para cuestionar el binomio rendimiento-calidad de la formación (aprendizaje) más allá de indicadores de resultados donde más importante que aprender es aprobar (Santos Guerra, 1999).

En este contexto, el presente trabajo plantea una propuesta centrada en la evaluación desde una orientación socioconstructivista del aprendizaje de las ciencias que pretende ayudar a superar algunas de las deficiencias señaladas y contribuir, en definitiva, a mejorar el aprendizaje y rendimiento de los estudiantes. Esta experiencia de innovación, realizada durante tres cursos académicos (2001-04), ha contado con la participación de alumnos de un primer curso de Universidad en la

asignatura Fundamentos Químicos de la Ingeniería (FQI), en la Escuela Técnica Superior del Medio Rural y Enología (ETSMRE) de la Universidad Politécnica de Valencia (UPV).

1.1. Espacio Europeo de Educación Superior: marco propicio para el cambio de cultura académica

Un antecedente importante en el camino al EEES reside en el Libro Blanco sobre la Educación y la Formación, surgido como síntesis de un amplio debate en el año 1995. En este documento, titulado “*Enseñar a Aprender: hacia la Sociedad del Conocimiento*”, como en muchos otros posteriores, se insiste en indicar que la clave de la reforma propuesta no está en la acomodación de los planes de estudio a una nueva estructura sino en un cambio de paradigma: pasar de una educación centrada en la enseñanza, a otra donde el centro de atención es el alumno y su aprendizaje, donde los docentes son facilitadores de ese aprendizaje.

La universidad del siglo XXI ha de profundizar en la misión educadora, priorizando el papel del estudiante y menos obsesionada con los planes de estudio, a los que se pretende dotar de una mayor flexibilidad. Este nuevo enfoque de la formación implica una revisión en profundidad de las metodologías de enseñanza-aprendizaje y en consecuencia de evaluación, con el fin de reflejar adecuadamente en las calificaciones el esfuerzo efectuado por los estudiantes (Docampo, 2001). Se trata, en suma, de una evaluación centrada en el desarrollo de competencias, que intenta superar el actual modelo educativo basado en la adquisición de conocimientos, en menor medida en aspectos procedimentales y claramente deficitario en la formación de actitudes (Ladino, 2001). Una evaluación que, vista globalmente, ha de tener en cuenta y valorar la educación como aventura personal, como un proceso formativo mediante el cual un individuo consigue desarrollar una visión propia del mundo a través del estudio de una disciplina científica, artística o tecnológica, proceso que ha de extenderse a lo largo de toda su vida (Delors et al., 1996).

En la actualidad este propósito contrasta con una visión generalizada de la evaluación reducida a su función calificadora. Si bien las calificaciones son o deberían ser una medida de los resultados de la enseñanza-aprendizaje, no lo son estrictamente de su calidad, pues están condicionadas por muchos factores que deben ser contemplados para incidir en una comprensión amplia de la evaluación y la calidad, atendiendo no sólo a resultados sino también a los procesos (Escudero, 1999; Santos Guerra, 1999; Boussada y De Ketele, 2005). Las actuales circunstancias invitan a los docentes a realizar una profunda revisión y reflexión sobre su práctica y, para los estudiantes, implica la llamada a una participación más activa y a una mayor corresponsabilidad en su formación. Pero, ¿cuáles son las ideas que

compartimos una gran mayoría de estudiantes y docentes acerca de la evaluación? ¿Qué aspectos deberían ser motivo de crítica y debate? En el apartado siguiente (1.2) se comentan algunas de estas cuestiones.

1.2. Cultura actual de la evaluación compartida por docentes y estudiantes

Existe un conjunto de creencias y concepciones que constituye una verdadera *cultura de la evaluación* generalizada y compartida por profesores y estudiantes, que puede caracterizarse por los siguientes aspectos:

- **La evaluación condiciona el proceso de enseñanza-aprendizaje, tanto desde la perspectiva de los docentes como de los alumnos.** Éstos comienzan el curso informándose sobre la proporción de suspensos en cursos anteriores y sobre qué mecanismos de adecuación deben desarrollar conforme a las exigencias de sus futuros profesores para lograr su supervivencia académica. La evaluación condiciona su trabajo, su forma de estudiar y sus relaciones con los docentes, con sus compañeros y con la institución. La selección y uso de unas estrategias de aprendizaje determinadas dependen en gran medida de sus percepciones sobre la eficacia académica de tales estrategias de cara a alcanzar los objetivos que marca el profesor (Alexander, Jetton, Kulikowich, y Woehler, 1994). De allí que la evaluación se considere como uno de los medios más efectivos de que dispone el profesor para influir en los estilos y modos de aprendizaje de los estudiantes (Crooks, 1988).

En una encuesta realizada en la Universidad Politécnica de Valencia el 75% el profesorado puso de manifiesto que los exámenes son la principal fuente de motivación de sus alumnos, reconociendo que constituyen prácticamente el único instrumento para acreditar en qué medida han aprendido (Llopis y Martínez, 2001; López et al., 2003). Citando otro ejemplo, en un estudio llevado a cabo en la Universidad Politécnica de Madrid, los estudiantes coincidieron en señalar la excesiva carga teórica de las asignaturas, aunada a un uso de los exámenes que convierte a sus estudios en una verdadera carrera de obstáculos. La opinión de un titulado dice textualmente que “los alumnos comienzan la carrera con inquietud por aprender y en menos de 6 meses, su intención pasa a ser simplemente aprobar exámenes” (González Tirado, 1994: 49).

Otros autores, como Sauleda y Martínez (1993: 259), llegan a afirmar que el alumno “trabaja en función de los exámenes”. Asimismo, como sostiene Santos Guerra (1999), resulta paradójico que se teorice sobre la importancia de la evaluación para la mejora del proceso de enseñanza, cuando por lo general se

repiten casi de forma mecánica unas mismas prácticas evaluativas. De este modo, pueden existir situaciones, tales como la coexistencia, en un mismo centro y asignatura de dos cursos con profesores diferentes, donde uno presenta el 80% de suspensos y el otro un porcentaje casi nulo.

- **Desvinculación sistemática de la evaluación y el proceso de enseñanza-aprendizaje.** Habitualmente suele sostenerse que el propósito de la evaluación no es sólo comprobar, sino mejorar el aprendizaje. Ello le confiere un carácter mediador (no finalista) y por lo tanto ejerce una función que se inserta y forma o debería formar parte fundamental del proyecto educativo. Sin embargo, en la práctica, se observa una desvinculación sistemática entre la evaluación y el proceso de enseñanza-aprendizaje (Rotger, 1990; Campanario, 1998). Así lo muestran los resultados de una encuesta llevada a cabo en la Universidad Politécnica de Valencia, donde el 60% del profesorado limita la evaluación a la aplicación de un examen final y el 74% manifiesta no hacer ninguna actividad evaluativa al terminar cada unidad didáctica o cada tema (Llopis y Martínez, 2001). No es de extrañar que “lo que aparece en los exámenes” se convierta para los estudiantes en sus objetivos de aprendizaje.

- **Coherencia entre el modelo de enseñanza-aprendizaje y los modos de evaluar.** La generalidad de la práctica habitual es consistente con un modelo de clase expositivo, caracterizado por el uso casi exclusivo de la lección magistral, la pasividad de los estudiantes y el énfasis en la memorización de conocimientos.

Existe una notoria contradicción entre lo que se pretende realizar o se manifiesta “querer” evaluar, como la capacidad para realizar análisis crítico, de relacionar y sintetizar, de desenvolverse con autonomía, etc. y lo que realmente se evalúa (Bol y Strage, 1996; Buendía y Olmedo, 2000; Pozo, 2001). Aunque las instituciones de formación de docentes hacen hincapié en la importancia de la evaluación cualitativa, la práctica evaluativa sigue instalada en modelos cuantitativos (Santos Guerra, 1999).

- **Existencia de fuertes preconcepciones en el profesorado respecto a la evaluación.** Desde el ámbito de la investigación se ha puesto de manifiesto que una gran mayoría de docentes asigna a la evaluación una función de instrumento de constatación final o simplemente acumulativa, además de considerarla casi con exclusividad limitada a los alumnos (Alonso, 1994).

Estas creencias están extendidas entre los docentes e incluso en la sociedad, generando ideas y comportamientos sobre la evaluación, que se apoyan entre sí y son coherentes con el modelo de enseñanza-aprendizaje en que se contextualizan (Alonso, 1994; Alonso, Gil y Torregrosa, 1996; Gabel, 1999). Algunas de estas concepciones espontáneas son: a) resulta fácil evaluar las materias científicas con objetividad y precisión (debido a la naturaleza misma de los conocimientos evaluados); b) el fracaso de un porcentaje significativo es inevitable en materias científicas; c) una prueba bien diseñada ha de ser discriminatoria y producir una distribución normal de notas, centrada en el 5; d) la función esencial de la evaluación es medir la capacidad y el aprovechamiento de los alumnos, asignándoles una puntuación que sirva de base objetiva para las promociones y selecciones, etc. (Alonso, 1994).

No obstante, diversos estudios demuestran hasta qué punto muchas de estas ideas son erróneas. Así, por ejemplo, diseños de investigación docimológica, solicitando la corrección de un mismo ejercicio de física o matemáticas, han mostrado que recibe puntuaciones muy diferentes cuando se califica por distintos profesores o por los mismos profesores en momentos diferentes (pasados tres o cuatro meses), lo que cuestiona la supuesta objetividad (López, Llopis, Llorens, Salinas, y Soler, 1983; Spears, 1984; De Ketele, 1985; Caverni, 1987; Alonso, 1994).

Las actuales prácticas evaluativas, tanto desde el punto de vista de la actuación de docentes y estudiantes como de la organización institucional universitaria, justifican en cierto modo la existencia de un amplio porcentaje de alumnos que no asisten a clase, la poca continuidad y dedicación en sus estudios, así como una deficiente percepción en los mismos estudiantes acerca de cuánto han aprendido o avanzado en la adquisición de determinadas competencias. En definitiva, se trata de prácticas que no favorecen un adecuado rendimiento académico y están lejos de ayudar a “aprender a aprender” (Delors et al., 1996). Por otra parte un modelo de enseñanza es algo más que un conjunto de elementos dispersos e intercambiables, siendo necesaria una coherencia de la que la evaluación forma parte (Crooks, 1988; Alonso, Gil y Torregrosa, 1992; Alonso, 1994). Puede afirmarse que la evaluación es el mecanismo que nos permite llegar a determinar hasta qué punto nuestra práctica docente es coherente con las finalidades y objetivos de nuestro proyecto formativo (Bol y Strage, 1996).

Los comentarios precedentes de ninguna manera llevan a obviar otras cuestiones, siendo de especial relevancia a la hora de intentar realizar cambios, tanto en las metodologías de enseñanza como en la evaluación, de problemas tales como la escasez o inadecuación de los recursos o la masificación. Éstos constituyen aspectos primordiales a la hora de intentar realizar con eficacia y calidad la tarea docente y se vienen señalando como algunos de los principales obstáculos para la

adaptación al sistema de créditos ECTS (Edwards, Ballester, Fuentes y Donderis, 2005). Como se verá más adelante, se ha utilizado una herramienta informática de apoyo para intentar paliar estos inconvenientes.

2. La evaluación integrada en el proceso de enseñanza-aprendizaje

Como alternativa a la problemática descrita en el apartado 1.2, durante tres años se ha ido ensayando y poniendo en práctica una propuesta de evaluación, coherente con una orientación socioconstructivista de la enseñanza y el aprendizaje de las ciencias, basada en la teoría del procesamiento de la información y del cambio conceptual (Posner, Strike, Hewson y Gerzog, 1982; Driver, 1986) y el aprendizaje estratégico (Pozo y Monereo, 1999); apoyada por investigaciones en aspectos como preconcepciones, resolución de problemas, trabajos prácticos, interacciones Ciencia-Tecnología-Sociedad-Medio Ambiente (CTSA) con aplicación de instrumentos tecnológicos (Burton, Holman, Pilling y Waddington, 1994; Kovac, 1999).

En cuanto a la propia incursión en el terreno del rendimiento académico y como ya se ha adelantado, hay que partir del hecho de que constituye un fenómeno complejo que se ve influido por múltiples factores, como los resultados académicos previos, el modo en como se eligen los estudios, el género, las condiciones familiares y socioeconómicas, etc. Respecto al rendimiento académico previo, una serie de estudios ha puesto de manifiesto que es un claro indicador de éxito en los estudios universitarios (Sánchez, 1995). No obstante, el rendimiento se centra en estos casos en las calificaciones, asumiéndolas como indicadores de aprendizaje. Heywood (2000) ha comprobado que el modelo de acceso a la universidad tiene una relación directa con los resultados obtenidos por los estudiantes y también con la tasa de abandono, lo que pone en primer plano de la discusión el dicho modelo, que no es, ni mucho menos, una cuestión baladí.

En cuanto a la evaluación, como ya se ha anticipado, se considera que ha de tener las siguientes características:

- Debe estar integrada en el proceso de enseñanza-aprendizaje, formando parte de este proceso y contribuyendo a mejorarlo. Permite al profesor, en cada etapa del proceso didáctico, tomar decisiones fundamentadas y ajustadas al proyecto educativo y facilita a los estudiantes regular su propio aprendizaje.
- Debe ser formativa, ya que su objetivo principal consiste en perfeccionar y enriquecer tanto el proceso como los resultados de la acción educativa. Insistimos en el papel de la evaluación formativa, dado su papel de refuerzo en la motivación de los alumnos, que se favorece cuando los mismos saben de

dónde parten y a dónde llegan, y están informados y se sienten protagonistas y partícipes de sus progresos (Hure et al., 2003).

- Debe ser continua, de manera que sus efectos no se produzcan sólo al final, al contrastar los resultados conseguidos, sino durante todo el proceso educativo, ya que, de otro modo, no podrían tomarse decisiones en el momento oportuno. Es necesario un diálogo permanente entre profesor y alumnos a lo largo del proceso didáctico para comprender la situación de aprendizaje en cada momento del proceso.
- Debe ser recurrente, en el sentido de reincidir a través de la realimentación sobre el desarrollo del proceso, perfeccionándolo de acuerdo con los resultados que se van alcanzando.
- Debe responder a criterios, lo cual supone formular previamente unos objetivos educacionales que orienten todo el proceso y permitan evaluar con rigor los resultados. Si no existen criterios previos, la evaluación no existe, ya que pierde todo punto de referencia y entra de lleno en el terreno de la indefinición, la ambigüedad o la subjetividad.
- Debe ser global, abarcando todos los factores que inciden en el proceso de enseñanza-aprendizaje, evaluando la evolución de cada estudiante pero incluyendo también todas las variables que intervienen en el desarrollo del proceso didáctico: las actividades de aprendizaje, el sistema de trabajo en el aula, los criterios de valoración, etc.
- Debe ser decisoria, los datos e informaciones debidamente tratados e interpretados han de facilitar la emisión de juicios de valor que, a su vez, propicien y fundamenten la toma de decisiones.
- Debe ser personalizada, es decir, realizada sobre la base del desarrollo de cada estudiante en particular. No se trata de ajustarse a unos patrones preestablecidos, sino de estudiar la evolución personal de los conocimientos y actitudes de los alumnos, su situación y estilo de aprendizaje y sus expectativas respecto al conocimiento científico, partiendo del diagnóstico inicial y controlando su progreso. Todo ello sin perjuicio de que el trabajo cooperativo sea también adecuadamente evaluado a partir de sus fines específicos y con los instrumentos adecuados.
- Consecuentemente, deben considerarse otros tipos de evaluación que potencian el aprendizaje, como es el caso de la evaluación entre pares, grupal, coevaluación y autoevaluación. En un meta-análisis de 122 estudios llevado a cabo por Johnson, Maruyama, Johnson, Nelson y Skon (1981) se pone de manifiesto la influencia del trabajo cooperativo y en grupo en la mejora del rendimiento académico respecto a prácticas evaluativas competitivas e individualistas.

3. Objetivos del presente estudio

El principal propósito de esta intervención ha sido implementar un conjunto de acciones para lograr que el mayor número posible de estudiantes pueda adquirir un nivel de competencias que les permita continuar los siguientes cursos con normalidad. Se trata de disminuir la elevada tasa de repeticiones o abandonos que se producen, particularmente en el primer año de universidad, procurando una mayor participación del estudiante en clase y su responsabilidad ante el aprendizaje. Se ha intentado favorecer una mayor motivación de los alumnos, hacerlos más partícipes y reflexivos sobre sus logros y dificultades, sus errores conceptuales, sus deficiencias metodológicas, etc., ayudándoles a introducir a lo largo del curso iniciativas de mejora (Kempa, 1991; Ben-Zvi y Hofstein, 1996).

En esta dirección, se parte indagando en qué medida las circunstancias con las que los estudiantes llegan a este primer curso influyen en su aprendizaje y rendimiento académico. Es decir, si aspectos tales como las calificaciones en la Prueba de Acceso a la Universidad (PAU), los conocimientos previos, las diferentes procedencias, la formación previa, entre otros, pueden considerarse factores predictivos del éxito académico. Algunos de esos aspectos podrían, en consecuencia, tenerse en cuenta de cara a un diseño y una planificación más adecuada del curso y de la actividad docente. Desde esta perspectiva, se plantean un conjunto de interrogantes que dan origen a la presente investigación, tales como:

- ¿en qué medida factores como la Selectividad, los estudios previos y algunos aspectos sociales influyen en el rendimiento académico de los alumnos al finalizar un primer curso de Química?
- ¿puede una evaluación diagnóstica inicial contribuir a una organización del curso y constituir un buen predictor del aprendizaje y el rendimiento ulterior de los estudiantes?
- ¿es posible mejorar el aprendizaje de nuestros alumnos y, en consecuencia, su rendimiento académico mediante la implementación de una evaluación formativa continua, coherente con una enseñanza de orientación socioconstructivista?
- ¿pueden presentar diferentes instrumentos de evaluación -específicamente las pruebas objetivas con cuestiones de elección múltiple respecto a cuestionarios abiertos o resolución de problemas- una capacidad predictiva similar en relación a la calificación?
- ¿existen algunos conceptos “nucleares” o “básicos” cuyo nivel de dificultad ha de ser tenido especialmente en cuenta por su incidencia en el aprendizaje y el rendimiento académico de los estudiantes?

A partir de estos interrogantes se han planteado las siguientes hipótesis:

Hipótesis 1

Una evaluación diagnóstica inicial en la que se tenga en cuenta un conjunto de factores (estudios previos, acceso universidad, examen de prerrequisitos) puede predecir con fiabilidad el desempeño y el rendimiento académico de los estudiantes así como ayudar a corregir las disfunciones observadas y orientar la intervención pedagógica en un primer curso de Química a nivel universitario.

Hipótesis 2

El diseño, selección y aplicación de intervenciones didácticas destinadas a instaurar un proceso de evaluación formativa continua en un curso de Química General pueden contribuir a la mejora del rendimiento y resultados del aprendizaje de los estudiantes.

Hipótesis 3

Desde el punto de vista de la calificación, los resultados obtenidos a partir de pruebas objetivas pueden ser similares a los proporcionados por otras actividades de evaluación, como cuestionarios abiertos o resolución de problemas.

Hipótesis 4

Es posible identificar un conjunto de contenidos conceptuales 'uncleares' o 'básicos' cuyo nivel de dificultad ha de ser tenido especialmente en cuenta por su incidencia en el aprendizaje y el rendimiento académico de los estudiantes.

Para tratar de dar respuesta a estas cuestiones se ha trabajado con una metodología que pasamos a describir en el apartado 4.

4. Metodología y diseño experimental

4.1. Aspectos generales y diseño global

En lo fundamental, se trata de un proyecto de investigación-acción que ha seguido la dinámica propuesta por autores como McKernan (1999) o Carr y Kemmis (1988), presentándose una propuesta a los estudiantes para que decidieran su participación voluntaria. Por investigación-acción se entiende el proceso mediante el

cual, en un área-problema determinada, el profesional en ejercicio lleva a cabo un estudio en primer lugar para definir con claridad el problema; en segundo lugar, para especificar un plan de acción, emprendiendo una evaluación para comprobar y establecer la efectividad de la acción tomada, observar los resultados y continuar con las modificaciones y realimentaciones necesarias.

En cuanto a contrastar las hipótesis anteriormente enunciadas, se han llevado a cabo las siguientes acciones concretas:

- Recogida de información y análisis estadístico para determinar en qué medida algunos aspectos pueden actuar como predictores del rendimiento académico.
- Estudio cualitativo de la información recogida en una primera fase de evaluación diagnóstica, teniendo en cuenta los aspectos precedentes y tomando decisiones respecto a las acciones necesarias para corregir las disfunciones observadas.
- Observación y seguimiento de todas las actividades de evaluación de los estudiantes y de la evolución de los aprendizajes individuales durante el curso académico.
- Estudio estadístico para probar que una mayor frecuencia en actividades de evaluación puede ser beneficiosa para el rendimiento académico.
- Estudio estadístico para comparar la eficacia de distintos instrumentos de evaluación, entre ellas exámenes y Cuestiones de Elección Múltiple (en adelante CEM).
- Estudio cualitativo y cuantitativo de la implementación del software París como instrumento para determinar las dificultades y errores conceptuales de los estudiantes y efectuar una adecuada selección de actividades en torno a determinados ejes o núcleos conceptuales, siempre con el propósito de consolidar un aprendizaje significativo.

4.2. Características de las actividades propuestas

Conforme a lo expuesto, las características generales de las actividades que se presentan en esta propuesta de evaluación han atendido a los siguientes criterios:

- Diseño y planificación inicial del curso (sujeta a futuros reajustes) con especial énfasis en la inclusión de contenidos conceptuales, procedimentales y actitudinales. Con una evaluación diagnóstica al inicio de curso por medio de la prueba *Química a la Entrada a la Universidad* (QEU).

- Uso de textos simplificados de apoyo, elaborados específicamente para su aplicación en el curso como unidades de aprendizaje independiente.
- Clases organizadas con una breve introducción de aproximadamente 10 minutos por parte del profesor y trabajo en grupos, ya sea en parejas o equipos de 4 ó 5 alumnos. En las mismas se ha tratado de trabajar armónicamente y de modo integrado la teoría, resolución de problemas y prácticas de laboratorio por parejas.
- Actividades de evaluación diversificadas, con aplicación de cuestionarios, pruebas tipo test, elaboración de informes de actividades experimentales, redacción de comentarios, exámenes parciales al finalizar cada unidad temática y globales, incluyendo algunas actividades de autorregulación. En relación a los exámenes, como sostiene Alonso (1994), no se trata de eliminarlos sino de convertirlos en ocasiones privilegiadas para desarrollar el aprendizaje, para aprender a aprender a través de la reflexión sobre los propios errores, retrocesos y avances que se van experimentando.
- Registro del seguimiento individualizado de los estudiantes a través del programa París, asistencia a tutorías, evaluación por pares, evaluación en grupo y autoevaluación.
- La calificación final no responde a una media o cálculo matemático de los registros parciales sino a un análisis integral de todos los momentos para determinar el grado de aprendizaje alcanzado por los estudiantes. En el curso 2002/03 se recogieron 34 registros (calificaciones) de las diferentes actividades realizadas, siendo de interés señalar que se evalúa cada unidad temática. Además, se lleva a cabo un seguimiento diario, con pruebas parciales del tipo test o CEM y abiertas (cuestionarios, resolución de problemas).

En todo el proceso se opta por un enfoque metodológico mixto, con procedimientos cualitativos y cuantitativos (Arnal, Del Rincón y Latorre, 1994). El análisis de los datos se realizó durante todo el curso lectivo, alternando en un proceso de interconexión y realimentación los momentos de recogida y análisis de información (Denzin y Lincoln, 1994). Se trata, además, de una investigación participativa, donde el investigador forma parte del contexto donde se lleva a cabo el estudio de campo, elabora registros de sus observaciones e interactúa en el contexto real del aula. Ésta se convierte, entonces, en un escenario propicio para comprender en cada caso específico el proceso de evaluación, para ir afrontando los problemas que se presentan, tratando de aportar soluciones.

4.3. El programa París

Es necesario destacar el papel de la tecnología de la información y comunicación (TIC), que ofrece posibilidades para diseñar una gran variedad de instrumentos de evaluación, así como contribuir en la organización del proceso evaluador, tanto en la recogida y gestión de la información como en su interpretación, facilitando la tarea docente (Cabero y Gisbert, 2002).

El programa utilizado (París) incluye opciones relacionadas con la gestión de ítems, de exámenes y de una base de datos de alumnos. Se trata de una herramienta muy potente desde el punto de vista de la facilidad de aplicación e interactividad. En el caso del presente estudio se utiliza como instrumento y al mismo tiempo como objeto propio de la investigación. Específicamente, el programa ha permitido estudiar y contrastar los niveles cognitivos de conocimiento, comprensión y aplicación de cada uno de los ítems seleccionados para el diseño de pruebas objetivas y del tipo abierto/problemas. Para una descripción más detallada del programa utilizado, consultar Apéndice.

A partir de la información recabada y comparando con otros estudios sobre preconcepciones y errores conceptuales (Pfund y Duit, 1994), se ha ido depurando el banco de ítems del programa, determinando los núcleos conceptuales y aspectos de mayor influencia en el aprendizaje y el rendimiento de los estudiantes.

5. Resultados

Durante los cursos 2000/01 y 2001/02 se realizaron pruebas piloto de la propuesta, lo que permitió ir ajustando las distintas fases de selección e implementación de las diferentes etapas. Los resultados que se exponen corresponden al curso 2002/03, en la que participaron 70 alumnos.

5.1. Resultados según los estudios previos de los estudiantes

Al comparar los estudios de procedencia, agrupamos los alumnos en PAU (estudiantes procedentes de bachillerato y COU que han participado en la Prueba de Acceso a la Universidad), en FP (los procedentes de módulos de Formación Profesional) y en U (los que han realizado estudios universitarios previos, finalizados o no). Coincidiendo con otras investigaciones (Hernández et al., 2002; González Tirado, 1994) se observa unas medias inferiores en las calificaciones del grupo procedente de FP respecto al procedente de PAU. La tabla 1 resume las procedencias de nuestros alumnos U, PAU y FP. El apartado “Sin datos” se refiere a los alumnos de los que no se tiene información, ya que, como se recordará, la participación en la experiencia ha sido voluntaria.

Muestra N = 70	Frecuencia	Porcentaje	Porcentaje válido
PAU	42	60,0	72,4
U	4	5,7	6,9
FP	12	17,1	20,7
Total	58	82,9	100,0
Sin datos	12	17,1	
Total	70	100,0	

Tabla 1. Procedencias de los alumnos en el curso

	Máximo	Media	Mediana	Mínimo	Desviación típica	Varianza
Selectividad	8,53	6,16	6,00	5,16	0,60	0,36

Tabla 2. Resumen de calificaciones en la Prueba de Acceso a la Universidad de los alumnos

En las figuras 1 y 2 se aprecian las distribuciones de las calificaciones obtenidas en el primer cuatrimestre y al finalizar el curso, en relación a las calificaciones obtenidas en la Prueba de Acceso a la Universidad.

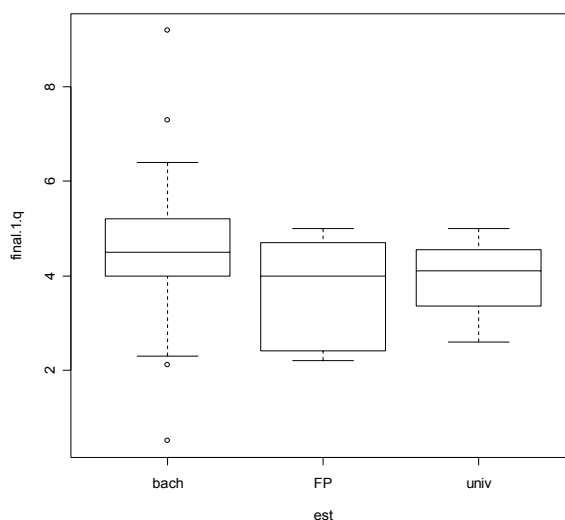


Figura 1. Estudios de procedencia frente a la calificación en el primer cuatrimestre

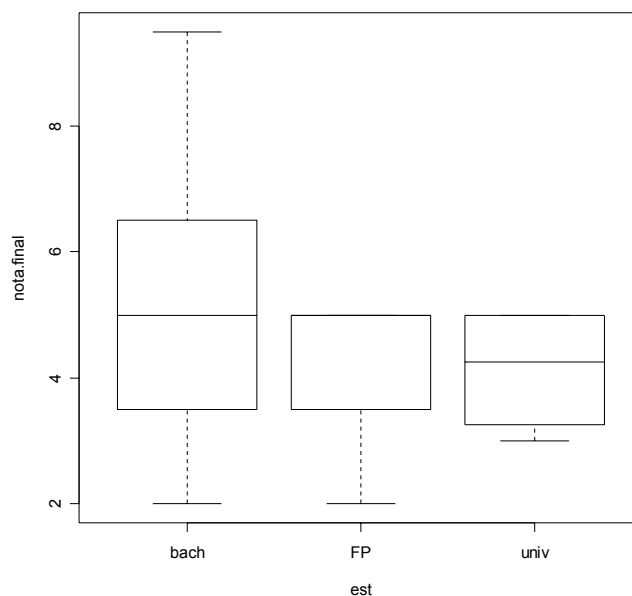


Figura 2. Estudios de procedencia frente a la calificación final del curso

Los resultados muestran que ningún estudiante procedente de FP y U obtuvo más de 5 puntos en la nota final, a diferencia de los de Bachillerato-COU y los universitarios. Estos datos, sumados a las dificultades detectadas en la evaluación diagnóstica al inicio del curso, confirman que las carencias de los alumnos procedentes de FP son superiores a las de los que proceden de Bachillerato-COU (y la necesidad de realizar acciones destinadas a dar una solución particularizada a este problema). También es significativo que haya abandonado el 75% de los alumnos procedentes de FP frente al 17,1% de los procedentes de PAU.

Por otra parte, y en función de información aportada por los mismos estudiantes, a través de entrevistas informales en pequeños grupos, muchos alumnos intentan sacar una buena calificación media en 1º y 2º de Bachillerato, matriculándose en asignaturas que, bien por ser minoritarias, por estar apoyadas por los profesores de esas áreas, o por resultarles a los alumnos de menor dificultad, les permiten, junto con la nota de acceso, aspirar a una media que les permita matricularse en la Facultad o Escuela deseada. De esta forma, se están produciendo importantes lagunas de formación en los alumnos que acceden a la universidad.

5.2. Relación entre la prueba diagnóstica inicial y el rendimiento académico

Las calificaciones obtenidas en la prueba diagnóstica denominada *Química a la Entrada a la Universidad* (QEU) han sido bastante bajas. Es necesario resaltar que estos resultados pueden presentar cierto sesgo, dado que los estudiantes sabían que los mismos no iban a ser considerados en su calificación final. No se ha observado correlación entre la prueba diagnóstica y las notas finales de cuatrimestre y curso, tal como se puede observar en la figura 3.

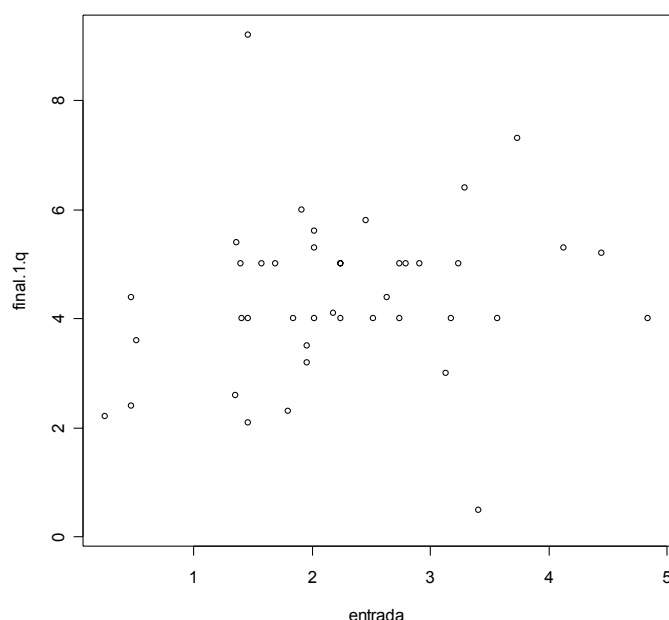


Figura 3. Relación entre la nota final del primer cuatrimestre y la calificación obtenida en la prueba QEU

El valor de la prueba QEU de principio de curso reside, sobre todo, en su carácter diagnóstico más que predictor de la futura calificación de los alumnos. Brinda la posibilidad, junto a las otras actividades iniciales, de que cada alumno pueda reconocer personalmente cuál es su situación frente a los contenidos y se puedan realizar tareas de refuerzo apropiadas.

Se puede afirmar en función de lo expuesto que no se ha hallado una relación clara entre la evaluación diagnóstica inicial y el rendimiento académico, expresado éste en la calificación al final del curso. En este sentido, ni las calificaciones en PAU ni la prueba QEU pueden predecir con fiabilidad el rendimiento académico de

los estudiantes. No obstante, sí se ha podido comprobar por observación que esta evaluación diagnóstica inicial puede ayudar a corregir las disfunciones observadas y reorientar la intervención pedagógica.

5.3. Influencia de la continuidad en las actividades de evaluación en el aprendizaje y el rendimiento académico

Los resultados obtenidos permiten afirmar que la mayor parte de los alumnos que han seguido el curso hasta el final, han obtenido una notable mejora en su aprendizaje, también traducida en sus calificaciones parciales o finales. De los 70 alumnos matriculados en el curso 2001/02, 23 abandonaron el curso, con lo que quedaron sin calificación final. De los que permanecieron, sólo dos obtuvieron un suspenso.

El resumen de las calificaciones obtenidas se presenta en la tabla 3, donde se muestran los resultados en el primer y segundo cuatrimestre y la calificación global del curso, una vez hechos los ajustes y correcciones necesarios para la calificación final.

ALUMNOS/AS	1C	% 1C	2C	% 2C	CURSO	% CURSO	
NO PRES.	17	24,3%	27	38,6%	23	32,9%	
SUSP.	13	18,6%	9	12,9%	2	2,9%	
APTOS 1C	20	28,6%			1	1,4%	
COMP 1C	20	28,6%			11	15,7%	
APTOS 2C			24	34,3%	3	4,3%	½ APTOS
COMP 2C			10	14,3%	2	2,9%	24,3%
APTOS CURSO¹					28	40,0%	
APTOS²	40	57,1%	34	48,6%	45	64,3%	

1C y 2C (1^{er} y 2^o cuatrimestre).

% 1C y % 2C (porcentaje de alumnos respecto al global en ambos cuatrimestres).

COMP. 1C y 2C (compensable 1er y 2^o cuatrimestre).

½ APTOS (los que sólo tienen que examinarse de medio curso, por tener la otra mitad calificada como apto o compensable).

Tabla 3. Resumen de las calificaciones obtenidas durante el curso

¹ Aptos curso frente a matriculados.

² Aptos frente a presentados.

Para cada uno de los grupos mostramos “no presentados”, “suspendidos”, “compensables” y “aptos” y porcentajes obtenidos. Como puede observarse, aprueba el curso completo el 40% de los alumnos y al menos una parte el 64,3%. Abandonan el modelo propuesto el 32,9%.

En el estudio de la correlación entre la media del número de actividades realizadas en el primer cuatrimestre (parciales) y la calificación obtenida se observa que (con independencia de que aprueben) los estudiantes que han realizado más actividades obtienen mejor resultado. En concreto, este resultado se obtuvo considerando dos muestras de estudiantes: un grupo que llevó a cabo 10 ó más actividades y otro donde este número fue inferior a 10, como se aprecia en la figura 4.

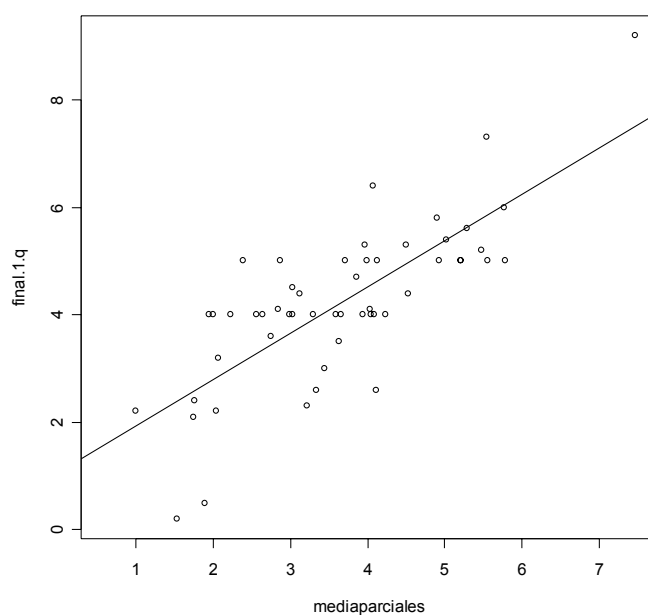


Figura 4. Media de actividades (parciales) realizadas frente a la calificación del primer cuatrimestre

Análogamente, en la figura 5 se presenta el esquema de la calificación final frente al número de actividades realizadas (t). De las 34 actividades calificadas en el curso, para esta prueba se considera los que han hecho ≥ 10 parciales ($t = 1$) y los que han hecho < 10 actividades ($t = 0$).

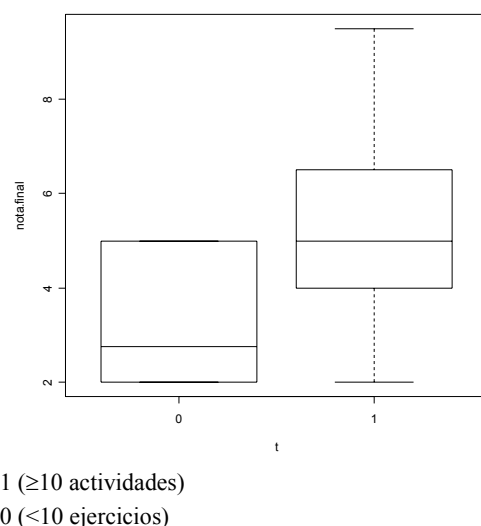


Figura 5. Calificación final frente a número de actividades “parciales” realizadas

Los que han hecho más de 10 “parciales” claramente obtienen mejor calificación, observándose diferencias en el número de aprobados al final de cuatrimestre y de curso.

5.4. Eficiencia y fiabilidad de las pruebas objetivas versus otros instrumentos de evaluación

Se ha obtenido una correlación elevada entre la valoración de las pruebas objetivas respecto de la calificación final. El análisis de regresión muestra que los que ejercicios abiertos (“man” en Fig. 6), es decir, cuestiones y resolución de problemas, no aportan información muy distinta respecto a esta valoración. Para comparar ambos instrumentos de evaluación se realizó un análisis de varianza de dos modelos: uno que comprende las pruebas *París* incluyendo actividades “abiertas” y otro que sólo considera pruebas objetivas de opción múltiple (CEM). Para esta comparación se diseña un modelo lineal en el que se agrupa, para el primer cuatrimestre, la media de actividades abiertas (mman) y la media de ejercicios París (mtest). Con unas gráficas de puntos se representan las medias “manual” y “París” frente a la calificación final del primer cuatrimestre. Los resultados muestran una elevada correlación entre ejercicios resueltos con el programa *París* y la nota final. El análisis de regresión pone de manifiesto que los resultados obtenidos (calificaciones) con ejercicios “abiertos” están significativamente correlacionados con los resultados (calificaciones) obtenidos a través del programa París.

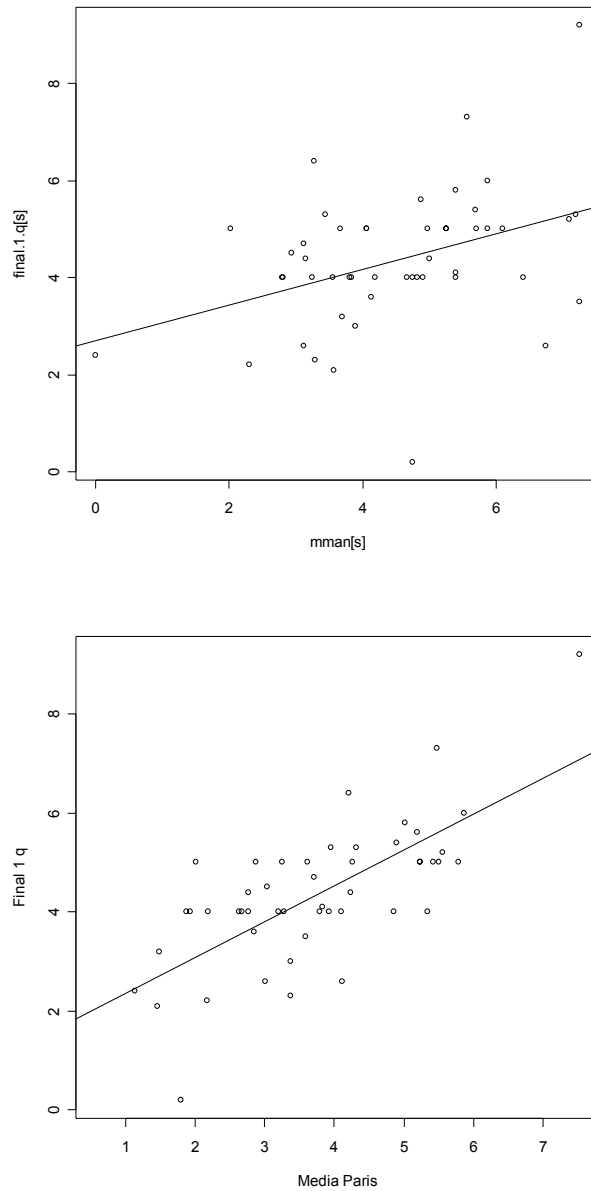


Figura 6. Modelo lineal comparando la media de actividades manuales (mman) y la media de CEM con el software Paris (mtest), a la izquierda. A la derecha, respecto a la calificación final (en ordenadas)

En el caso de las CEM implementadas con el programa *París* se obtiene una mayor pendiente (1,62), ordenada en el origen 1,63, observándose además una mayor agrupación en los valores. Estos datos confirman la tercera hipótesis de que los resultados obtenidos a partir de pruebas objetivas pueden ser similares a los proporcionados por otras actividades de evaluación, como cuestionarios abiertos o resolución de problemas.

En la tabla 4 se observa no sólo las buenas correlaciones con las notas finales, sino la elevadísima correlación entre parciales y parciales manuales, lo que confirma que una de ellas (*París*) puede ser suficiente, sin necesidad de duplicar pruebas.

		sumatorio parciales	sumatorio parc. Manual (10)	final cuatri- mestre 1	CALIFI- CACIÓN FINAL
Sumatorio parciales	Correlación de Pearson	1	0,940(**)	0,751(**)	0,720(**)
	Sig. (bilateral)	0	0,000	0,000	0,000
	N	64	62	51	46
Sumatorio parc. manual (10)	Correlación de Pearson	0,940(**)	1	0,706(**)	0,608(**)
	Sig. (bilateral)	0,000	0	0,000	0,000
	N	62	62	51	46
final cuatrimestre 1	Correlación de Pearson	0,751(**)	0,706(**)	1	0,804(**)
	Sig. (bilateral)	0,000	0,000	0	0,000
	N	51	51	53	45
CALIFICACIÓN FINAL	Correlación de Pearson	0,720(**)	0,608(**)	0,804(**)	1
	Sig. (bilateral)	0,000	0,000	0,000	0
	N	46	46	45	47

** La correlación es significativa al nivel 0,01 (bilateral).

Tabla 4. Correlaciones de totales *París* y manual frente a calificaciones finales

Este modelo de evaluación formativa continua exige un gran esfuerzo por parte del profesor, no sólo por el seguimiento diario sino sobre todo por las tareas de corrección de informes y comentarios, desde el planteamiento de la realización de pruebas parciales a la finalización de cada unidad temática. De ahí la importancia de confeccionar un banco de ítems que permita al profesor, una vez en funcionamiento la evaluación continua, afrontar el problema de la corrección de un número elevado de exámenes en cursos masificados. Ello le posibilitaría atender mejor otras tareas contribuyendo a optimizar la gestión del tiempo.

5.5. Identificación de contenidos conceptuales nucleares con aplicación del software París

La parte más significativa y compleja de la actual propuesta reside en la selección y depuración de cada uno de los ítems en la base de datos. Para ello se ha realizado un estudio cualitativo centrado en los contenidos conceptuales, sumado a los análisis aportados por los índices de discriminación y facilidad del programa y los resultados aportados por la investigación en didáctica y enseñanza de la química. Para cada uno de los contenidos se han considerado tres niveles: conocimiento, comprensión y aplicación. Se ha traducido los dos bancos de ítems de la *American Chemical Society*, realizando una rigurosa selección de esta información, junto a otras bases de datos, como los de *Examination Boards* británicos y del grupo *Chapam* francés. Además, se ha estudiado la posible correlación entre núcleos “más problemáticos” o de mayor dificultad -de acuerdo a la investigación existente al respecto- para los estudiantes con la finalidad de diseñar actividades orientadas específicamente, tanto al propio desarrollo del curso como en la elaboración de materiales de autoaprendizaje. En las tablas 5 y 6 se detallan, como ejemplo, los valores calculados para cada uno de los temas mediante regresión lineal.

Modelo	Núcleos conceptuales	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	3,970E-02	0,615		0,065	0,949
	estequiometría	0,263	0,111	0,356	2,370	0,029
	ea+sp	8,935E-04	0,135	0,001	0,007	0,995
	enlace	0,195	0,086	0,252	2,251	0,036
	termo	0,152	0,110	0,242	1,372	0,186
	equilibrio	7,517E-02	0,083	0,127	0,902	0,378
	ácido-base	0,165	0,098	0,208	1,684	0,109
	complejos	-0,103	0,100	-0,116	-1,026	0,318
	solubilidad	-0,078	0,095	-0,110	-0,822	0,421
	redox	0,297	0,101	0,337	2,951	0,008

Variable dependiente: final cuatrimestre 1, ea + sp: estructura atómica y sistema periódico

Tabla 5. Cálculo de coeficientes de regresión lineal, parciales frente a Final primer cuatrimestre

Modelo	Núcleos conceptuales	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	-0,564	1,067		-0,529	0,604
	estequiometría	0,271	0,167	0,279	1,625	0,124
	ea+sp	4,431E-02	0,232	0,045	0,191	0,851
	enlace	0,106	0,139	0,099	0,762	0,457
	termo	0,424	0,167	0,506	2,534	0,022
	equilibrio	-0,161	0,130	-0,202	-1,232	0,236
	ácido-base	0,164	0,152	0,164	1,080	0,296
	complejos	0,107	0,153	0,093	0,701	0,493
	solubilidad	-0,082	0,144	-0,088	-0,573	0,575
	redox	0,386	0,161	0,326	2,399	0,029

Variable dependiente: NOTA FINAL, ea + sp: estructura atómica y sistema periódico

Tabla 6. Cálculo de coeficientes de regresión lineal, parciales frente a calificación final

Como puede observarse en Tablas 5 y 6, existen unos núcleos conceptuales de mayor incidencia en la calificación al finalizar el primer cuatrimestre: redox, estequiometría y enlace y ácido-base, termo, redox y estequiometría, respecto a la calificación final.

En la tabla 7 se presentan las correlaciones entre los distintos núcleos conceptuales y las calificaciones finales de cuatrimestre y final, siendo los más importantes en cuanto a su carácter predictor: estructura atómica, estequiometría, termo (primer cuatrimestre) y estructura atómica, termo, estequiometría (calificación final).

Núcleos conceptuales	Correlación	final cuatrimestre 1	CALIFICACIÓN FINAL
estequiometría	Correlación de Pearson	0,655(**)	0,590(**)
	Sig. (bilateral)	0,000	0,000
	N	50	43
ea+sp	Correlación de Pearson	0,665(**)	0,608(**)
	Sig. (bilateral)	0,000	0,000
	N	47	40
Enlace	Correlación de Pearson	0,381(**)	0,242
	Sig. (bilateral)	0,010	0,113
	N	45	44
Termo	Correlación de Pearson	0,621(**)	0,606(**)
	Sig. (bilateral)	0,000	0,000
	N	45	44
equilibrio	Correlación de Pearson	0,448(**)	0,366(*)
	Sig. (bilateral)	0,002	0,015
	N	45	44
ácido-base	Correlación de Pearson	0,415(*)	0,532(**)
	Sig. (bilateral)	0,013	0,002
	N	35	32
complejos	Correlación de Pearson	0,193	0,360(*)
	Sig. (bilateral)	0,267	0,043
	N	35	32
solubilidad	Correlación de Pearson	0,410(*)	0,428(*)
	Sig. (bilateral)	0,014	0,015
	N	35	32
Redox (enero) 10	Correlación de Pearson	0,538(**)	0,315(*)
	Sig. (bilateral)	0,000	0,040
	N	51	43
final cuatrimestre_1	Correlación de Pearson	1,000	0,804(**)
	Sig. (bilateral)	0,000	0,000
	N	53	45
NOTA FINAL	Correlación de Pearson	0,804(**)	1,000
	Sig. (bilateral)	0,000	0,000
	N	45	47

** La correlación es significativa al nivel 0,01 (bilateral).

* La correlación es significativa al nivel 0,05 (bilateral).

Tabla 7. Correlaciones entre núcleos conceptuales y la calificación final

Los análisis cualitativos realizados corroboran estos resultados, confirmando nuestra cuarta hipótesis de que es posible identificar un conjunto de contenidos conceptuales ‘nucleares’ o ‘básicos’ cuyo nivel de dificultad ha de ser tenido especialmente en cuenta por su incidencia en el aprendizaje y el rendimiento académico de los estudiantes.

6. Conclusiones y comentarios

Se ha presentado, a grandes rasgos, los resultados de una investigación sobre una propuesta de innovación en un primer curso de química a nivel universitario, en la que se ha intervenido con un sistema de evaluación continua y formativa, integrada en el proceso de enseñanza-aprendizaje.

Se ha observado la existencia de diferencias de nivel entre los alumnos que provienen de Bachillerato-COU y los de FP, siendo estos últimos los que presentan mayor tasa de abandono (75% frente al 17,1% de los procedentes de PAU). Estos resultados, coincidentes con otras investigaciones, indican la necesidad de replantear las condiciones de acceso de los alumnos que proceden de FP o de adoptar medidas para compensar este déficit en la formación previa.

Respecto a la primera hipótesis, no se ha encontrado una correlación significativa entre la calificación de la Prueba de Acceso a la Universidad y la calificación final obtenida en el curso, lo que contradice la opinión generalizada al respecto. Tampoco se ha encontrado una relación clara entre la evaluación diagnóstica inicial y el rendimiento académico, expresado en la calificación final del curso. En este sentido, las actividades realizadas (como la prueba *Química a la Entrada a la Universidad*, por ejemplo) no son predictivas respecto a la calificación final, pero son muy importantes para orientar sobre el nivel de competencias iniciales de los estudiantes y, a partir de ahí, organizar de modo más efectivo las actividades del curso.

Tanto la información recabada con el programa *París* como los análisis cualitativos han permitido poner de manifiesto cuáles son los núcleos conceptuales que resultan más problemáticos para los estudiantes y aquéllos que les resultan más sencillos; considerados en la implementación de las actividades iniciales de nivelación y refuerzo. El desconocimiento, por ejemplo, de la estequiometría, conduce a errores posteriores en casi todos los otros temas.

Se ha comprobado que el modelo y el conjunto de actividades de enseñanza-aprendizaje propuestos dentro de este proyecto de evaluación formativa continua, con los matices apuntados, ayuda a mejorar el rendimiento académico de los alumnos. El mayor nivel de participación se ha traducido en una mayor asistencia a clase, una menor tasa de abandono, en el número de actividades realizadas así como en el desarrollo y el trabajo colaborativo entre los estudiantes y el profesor.

En el curso 2001-2002 sólo 2 estudiantes, de los que siguieron el modelo propuesto, no aprobaron. Aprobó todo el curso el 40% de los alumnos y al menos el primer cuatrimestre, el 64,3%. En la fase piloto del curso previo, los resultados fueron similares: de 69 alumnos, fueron abandonando progresivamente la propuesta 17 estudiantes y no aprobaron el curso sólo 9. Comparativamente y para dar una idea clara del grado de mejora obtenido, según un estudio realizado con una cohorte de estudiantes en el período 1995-2000 en la Universidad Politécnica de Valencia, Química I es una asignatura señalada como de alto grado de dificultad, con un 29% de aprobados en la convocatoria ordinaria.

En relación a la comparación entre las pruebas objetivas (CEM) y otros tipos de instrumentos de evaluación, los resultados han mostrado una elevada correlación entre CEM y la calificación final. El análisis de regresión pone de manifiesto que los ejercicios “abiertos” aportan, comparativamente, la misma información que las pruebas objetivas respecto a la calificación, confirmando la tercera hipótesis. Ello permite sostener que las a veces despreciadas pruebas objetivas pueden ser instrumentos útiles, siempre que respondan a un diseño y selección adecuados. Este modelo de evaluación continua exige un gran esfuerzo por parte del profesor por lo que la confección de un banco de ítems puede ayudar en gran medida tanto en la tarea de elaboración de actividades de evaluación como en la gestión de todo el proceso.

La versatilidad del programa *París* permite preparar bancos de actividades de cualquier materia y, aunque esto supone un gran esfuerzo inicial, puede ser de gran ayuda para las tareas de evaluación, tanto para estudiantes como para los profesores. Mediante la información que suministra sobre índices de facilidad y de discriminación de las pruebas o actividades de evaluación elaboradas, puede facilitar el diagnóstico y reajuste del proceso de enseñanza-aprendizaje para cada curso en particular, y permite trasladar una experiencia educativa para que pueda ser reproducida en otros contextos

Su aplicación ha sido de mucha utilidad para corroborar la cuarta hipótesis, identificando un conjunto de contenidos conceptuales “nucleares” o “básicos” cuyo nivel de dificultad ha de ser tenido especialmente en cuenta por su incidencia en el aprendizaje y el rendimiento académico de los estudiantes. Los conocimientos sobre estequiometría (con su repercusión en la interpretación y relaciones cuantitativas en todo tipo de reacciones) y termoquímica y equilibrio en general (que precisa del manejo de unidades de energía y relaciones estequiométricas y de equilibrio) muestran un papel claramente predictivo del rendimiento en la asignatura. Ello, por otra parte, es fácil de entender si tiene en cuenta que dichos conocimientos son la base para la comprensión del equilibrio químico; concepto que, a su vez, constituye el hilo conductor del resto de los temas del curso lo que obliga a considerar la necesidad de reforzar los conocimientos previos sobre dichos aspectos.

Organizar la enseñanza para un aprendizaje significativo requiere que el profesorado sepa cómo detectar las dificultades del alumnado y los obstáculos con que se encuentra a lo largo de su aprendizaje. Para que los estudiantes puedan aprender se requiere que ellos mismos reconozcan estas dificultades y obstáculos, para así poder pedir y encontrar los medios para superarlos eficazmente. Ello exige una nueva cultura de la evaluación, en la que las actividades evaluativas lleguen a ser al mismo tiempo actividades de enseñanza-aprendizaje, integradas en todo el proceso formativo.

El reto para los docentes está en cómo conseguir que los estudiantes vayan adquiriendo la habilidad de autorregularse de manera efectiva. Para conseguirlo los estudiantes necesitan sentir como propios los objetivos y los criterios de evaluación. Esta nueva cultura de la evaluación también requiere cambiar el entorno de aprendizaje, cuidar el ambiente del grupo y unos valores que faciliten la verbalización de las ideas y de las formas de trabajo, el intercambio de opiniones y puntos de vista, el respeto mutuo y la elaboración de propuestas consensuadas. Por tanto, avanzar en armonía hacia el EEES supone trabajar por una nueva cultura de la evaluación que sea al mismo tiempo una cultura de calidad, presente en el quehacer diario y en el trabajo coherente y cooperativo de estudiantes y docentes.

Bibliografía

- Alexander, P. A., Jetton, T. L., Kulikowich, J. M. y Woehler, C. (1994). Contrasting instructional and structural importance: The seductive effects of teacher questions. *Journal of Reading Behaviour*, 26: 19-45.
- Alonso, M. (1994). *La evaluación en la enseñanza de la Física como instrumento de aprendizaje*. Tesis Doctoral no publicada. Departament de Didàctica de les Ciències Experimentals i Socials. Universitat de València.
- Alonso, M., Gil, D. y Martínez Torregrosa, J. (1992). Concepciones espontáneas de los profesores de Ciencias sobre la evaluación. Obstáculos a superar y propuestas de replanteamiento. *Revista de Enseñanza de la Física*, 5(2): 18-38.
- Alonso, M.; Gil, D. y Martínez Torregrosa, J. (1996). Evaluar no es calificar. La evaluación y la calificación en una enseñanza constructivista de las Ciencias. *Educación en la Escuela*, 30: 15-26.
- Arnal J., Del Rincón D. y Latorre A. (1994). *Investigación educativa. Fundamentos y metodología*. Barcelona: Labor.

- Attali, J. (Coord.). (1998). Pour un modèle européen d'enseignement supérieur. Rapport de la Commission. <http://www.lemonde.fr/dossiers/attali/> [ref. 13 abril 2004].
- Ben-Zvi, R. y Hofstein, A. (1996). Strategies for remediating learning difficulties in chemistry. En Treagust, D. F., Duit, R. y Fraser, B. J. (eds.). *Improving teaching and learning in science and mathematics*. New York: Teachers College Press. 109-119.
- Bol, L. y Strage, A. (1996). The contradiction between teachers' instructional goals and their assessment practices in high school biology courses. *Science Education* 80: 145-163.
- Bricall, J. M. (2000). *Universidad 2000*. Madrid: CRUE.
- Boussada, H. y De Ketele, J. M. (2005). *L'évaluation comme indicateur de la qualité de la formation universitaire: étude empirique dans le cadre du mastère de didactiques spécialisés*. 22^e Congrès de l'Association Internationale de Pédagogie Universitaire (AIPU). Genève, septembre de 2005.
- Buendía Eisman, L. y Olmedo Moreno, E. M. (2000). Estrategias de aprendizaje y procesos de evaluación en la educación universitaria. *Bordón*, 52(2): 151-163.
- Burton, W. G., Holman, J. S., Pilling, G. M. y Waddington, D. J. (1994). *Advanced Chemistry Salters: Chemical storylines, chemical ideas, activities and assessment pack. Teacher's guide*. Oxford: Heinemann.
- Cabero, J. y Gisbert, M. (Dirs). (2002). *Materiales formativos multimedia en la red. Guía práctica para su diseño*. Sevilla: SAV de la Universidad de Sevilla.
- Campanario, J. M. (1998). Preguntas y respuestas sobre la evaluación de los alumnos en la enseñanza de las Ciencias. *Tarbiya: Revista de Investigación e Innovación Educativa*, 19: 69-84.
- Campanario, J. M. (2002). Asalto al castillo: ¿a qué esperamos para abordar en serio la formación didáctica de los profesores universitarios de Ciencias? *Enseñanza de las Ciencias*, 20(2): 315-325.
- Carr, W. y Kemmis, S. (1988). *Teoría crítica de la enseñanza*. Barcelona: Martínez Roca.
- Caverni, J. P. (1987). Knowledge acquisition assessment by experts: Effects and models of the cognitive functioning of evaluators. *European Journal of Psychology of Education*, 2: 119-131.
- Comission on Higher Education MSA (1997). Framework for Outcomes Assessment. Middle States Association.

- Comisión de las Comunidades Europeas. (1995). *Libro Blanco sobre la educación y la formación - Enseñar y aprender - Hacia la sociedad cognitiva*. Monográfico COM (95) 590, noviembre de 1995.
http://europa.eu.int/comm/off/white/index_es.htm [Ref. 20 junio 2004].
- Comisión de las Comunidades Europeas. (2003). *El papel de las Universidades en la Europa del conocimiento*. Bruselas. Com (2003) 58 final.
http://www.uah.es/universidad/espacio_europeo/documentos/papel_las%20univ_europa_conocimiento%20febrero_2003.pdf [Ref. 13 noviembre 2004].
- Consejo de Coordinación Universitaria. (2003). *Plan Nacional de Evaluación de la Calidad de las Universidades. Informe Global (1996-2000)*. Madrid: Secretaría General Técnica.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58: 438-481.
- Dearing, R. (1997). *Higher education in learning society*.
<http://www.leeds.ac.uk/educol/ncihe/> [Ref. 20 abril 2005].
- De Ketele, J. M. (1985). *Docimologie. Introduction aux Concepts et aux Pratiques*. Louvain la Neuve: Ed. Cabay.
- Delors, J. (Coord.). (1996). *Informe Delors. La educación encierra un tesoro*. Madrid: Unesco-Santillana.
- Denzin, N. K. y Lincoln, Y. S. (Eds.). (1994). *Handbook of qualitative research*. London: Sage Publications.
- Driver, R. (1986). Psicología cognoscitiva y esquemas conceptuales de los alumnos. *Enseñanza de las Ciencias*, 4(1): 3-16.
- Docampo, D. (2001). *Educación centrada en el aprendizaje*. Conferencia pronunciada na Universidade de Granada en Novembro de 2001. Actas das Xornadas sobre Tutorías e Orientación.
- Edwards Schachter, M., Ballester Sarrias, E., Fuentes Durá, P. y Donderis Quíles, V. (2005). *Perceptions et avis du professorat espagnol sur le processus de convergence. XXIIe Congrès de l'AIPU*. Ginebra. Septiembre de 2005.
- Escudero, T. (1999). Indicadores del rendimiento académico: una experiencia en la Universidad de Zaragoza. En Vidal, J. (coord.). *Indicadores en la Universidad: información y decisiones*. Madrid: Consejo de Universidades, MEC. 251-262.
- Fabra, M. L. y Domènech, M. (2001). *Hablar y escuchar. Relatos de profesor@s y estudiant@s*. Barcelona: Papeles de Pedagogía/ Paidós.

- Gabel, D. (1999). Teaching and Learning Chemistry Education Research: a look to the future. *Journal of Chemical Education*. 76 (4): 548-555.
- Gabel, D. L. y Bunce, D. M. (1994). Research on problem solving: Chemistry. En Gabel, D. L. *Handbook of research on science teaching and learning*. New York: Macmillan Publishing Company. 301-326.
- González Tirado, R. M. (1994). *Análisis de los factores de proceso y otros factores de influencia que condicionan la calidad de las enseñanzas universitarias y el rendimiento (Estudio Piloto)*. Madrid: ICE de la Universidad Politécnica de Madrid.
- Hernández Armenteros, J. (2002). *Información académica, productiva y financiera de las universidades públicas españolas. Año 2000. Indicadores universitarios (curso académico 2000-2001)*. Madrid: Conferencia de Rectores de las Universidades Españolas.
- Hernández Marco, J., Bará Temes, J., Córdoba Pérez, J. F., De Luis, R. y Martín Castro, P. (2002). *Evaluación transversal del rendimiento académico de las ingenierías técnicas*. Plan Nacional de Evaluación de la Calidad de las Universidades. Madrid: Ministerio de Educación y Ciencia.
- Heywood, J. (2000). *Assessment in Higher Education*. London: Jessica Kingsley Pub.
- Hure, M.E., Torres, C.M., De Vicente Pérez, S., Galli, N., Beccarece, S. y Marini, M. (2003). Evaluación continua en química general e inorgánica. *Anuario Latinoamericano de Educación Química*. 2002-2003. Nº. XVI: 282-286.
- Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D. y Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin*, 89: 47-62.
- Kempa, R. (1991). Students' learning difficulties in science. Causes and possible remedies. *Revista de Enseñanza de las Ciencias*, 9(2): 119-128.
- Kovac, J. (1999). Student active learning methods in general chemistry. *Journal of Chemical Education*, 76(1): 120-124.
- Ladino Ospina, Y. (2001). Sobre la evaluación de competencias. VI Congreso Internacional sobre Investigación en la Didáctica de las Ciencias. Retos de la enseñanza de las Ciencias en el Siglo XXI. *Revista Enseñanza de las Ciencias* Nº extra. Comunicaciones. Tomo 1, 233-234.
- Llopis Castelló, R. y Martínez Aparisi, A. (2001). *Esfuerzos del profesor en la organización del aprendizaje de sus alumnos en la UPV*. I Jornadas de Innovación Educativa. ICE de la Universidad Politécnica de Valencia, septiembre de 2001.

- López, S., Calviño, F., De Corral, F., Llopis, R., Martí, R., Martín, P., Martínez, A. y Armengol, J. (2003). *Cómo se organiza el aprendizaje y su evaluación en la UPC: nuestra aproximación a través de una encuesta*. IX Congreso Universitario de Innovación Educativa en las Enseñanzas Técnicas. Universidad Politécnica de Cataluña, Vilanova i la Geltrú, julio de 2003.
- López, N., Llopis, R., Llorens, J. A., Salinas, B. y Soler, J. (1983). Análisis de dos modelos evaluativos referidos a la Química de COU y selectividad. *Enseñanza de las Ciencias*, 1(1): 21-25.
- Marchesi, Á. (2003). Indicadores de la educación en España y cambio educativo. *Revista de Educación*, 330 (enero-abril). Monográfico Reflexiones sobre política educativa, 13-34.
- McKernan, J. (1999). *Investigación-acción y curriculum*. Madrid: Morata.
- Pfund, H. y Duit, R. (1994). *Bibliography – Students' Alternative Framework and Science Education* (4th Edition). Kiel: Institute for Science Education.
- Posner, G. J., Strike, K. A., Hewson, P. V. y Gerzog W. A. (1982). Accommodation of a scientific conception: Towards a theory of conceptual change. *Science Education*, 66: 211-227.
- Pozo, J. I. (2001). Curso Universitario: ¿Por qué no aprenden los alumnos universitarios lo que se les quiere enseñar? Cededuis *Revista Docencia Universitaria*, 2(2), <http://www.uv.es/~arbelaez/v2n212curso.htm> [Ref. 18 septiembre 2005].
- Pozo, J. I. y Monereo, C. (Coord.). (1999). *El aprendizaje estratégico*. Madrid: Santillana.
- Rodríguez Espinar, S. (2003). *La problemática del rendimiento en la transición Secundaria-Universidad*. Actas X Jornadas sobre Acceso a la Universidad. Universidad de Sevilla.
- Rotger Amengual, B. (1990). *Evaluación formativa. Educación y Futuro. Monografías para la Reforma*. Madrid: Editorial Cincel.
- Sánchez Gómez, M. C. (1995). *Determinantes del rendimiento académico en la Universidad de Salamanca*. Tesis doctoral. Departamento Didáctica, organización y Métodos de Investigación. Universidad de Salamanca.
- Santos Guerra, M. A. (1999). 20 Paradojas de la evaluación del alumnado en la Universidad española. *Revista Electrónica Interuniversitaria de Formación del Profesorado* 2(1). Disponible en <http://www.uva.es/aufop/publica/revelfop/99-v2n1.htm> [Ref. 20 junio 2005].

- Sauleda Pares, N. y Martínez Ruiz, M. (1993). Propensiones en la evaluación de las actividades prácticas propias de las ciencias experimentales: Un escenario para el cambio. *Revista Interuniversitaria de Formación del Profesorado* 18: 255-271.
- Smith, T. M. (1999). Comparaciones internacionales sobre educación terciaria. En Vidal, J. (coord.). *Indicadores en la Universidad: información y decisiones*. Plan Nacional de Evaluación de la Calidad de las Universidades. Madrid: Consejo de Universidades, MEC. 31-50.
- Spears, M. G. (1984). Sex bias in science teachers' ratings of work and pupils characteristics. *European Journal of Science Education*, 6: 369-377.
- Yorke, M. (1999). *Leaving early: Undergraduate non-completion in higher education*. London: Falmer Press.

Apéndice. Posibilidades del programa París

- **Gestión de Ítems.** Permite construir múltiples bancos de ítems organizados por temas, brindando información acerca de índices de facilidad y discriminación de los mismos, fecha de creación y utilización, palabras claves y cualidades que describe cada uno de los ítems. Se pueden introducir ecuaciones, comentarios, imágenes, audio o video. Es compatible con otras aplicaciones y permite conexiones a otros programas (MathLab, Derive, etc.). En esta aplicación se cuenta en la actualidad con un banco de 3.000 actividades diferentes para el área de Química que se pueden seleccionar y también modificar. Están elaboradas teniendo en cuenta la traducción y adaptación de los bancos de ítems *General Chemistry* y *High School Chemistry* de la *American Chemical Society*. También se han incorporado actividades provenientes de aportaciones de la investigación en didáctica de la Química, adaptaciones de actividades del proyecto *Salter's*, ideas surgidas a partir de colaboraciones de docentes y estudiantes, etc. (Burton et al., 1994).
- **Gestión de Exámenes.** Esta opción, aunque calificada de este modo por sus aplicaciones iniciales, se refiere a la manipulación de la información contenida en las actividades que pueden diseñarse. El programa no se reduce a la aplicación de pruebas objetivas, sino que permite el diseño de una amplia variedad de actividades de evaluación (desde tests de respuesta corta a respuesta abierta, clasificación de elementos, mapas conceptuales, etc.). Se pueden clasificar y manipular según diferentes criterios, como fecha de última utilización, dificultad, temas, autores, elección aleatoria, etc. Se pueden imprimir, exportar a Word o a otra aplicación, situar en una página web o guardarse para usarse a través de un módulo denominado *ParisExa*. Dicho módulo incluye múltiples opciones para la introducción de los datos personales del alumno, pausa y reanudación inmediata del examen o actividad por el profesor, control de tiempo transcurrido para cada alumno, etc. Posibilita la corrección automática de exámenes tipo test a partir de los resultados obtenidos mediante lectora óptica o corrección directa. Las calificaciones obtenidas se asignan a los alumnos de manera directa en los archivos correspondientes y se pueden pasar a otros formatos para ser utilizadas desde otros programas, permitiendo además la creación de diferentes tipos de informes.
- **Gestión de una base de datos de alumnos.** Puede colocarse información personal junto a calificaciones, pudiendo estudiar la correlación entre ambas. Permite la creación de fórmulas para calcular nuevas columnas de valores (como una hoja de cálculo) y proporciona opciones para ordenar y seleccionar información de los alumnos según diversos criterios y alternativas, así como para exportar e importar datos a/desde otros formatos.