

## **Evaluación entre compañeros: Cómo lo hacemos en nuestros cursos de programación de ordenadores**

Pablo del Canto, Isabel Gallego, José Manuel López, Javier Mora, Angélica Reyes, Eva Rodríguez, Kanapathipillai Sanjeevan, Eduard Santamaría, Miguel Valero

*Departament d' Arquitectura de Computadors,  
Escola Politècnica Superior de Castelldefels.  
Universitat Politècnica de Catalunya  
Avinguda del Canal Olímpic, 15. 08860 Castelldefels. Telf. 934137000*

### **1. Introducción**

Cuando hablamos de evaluación entre compañeros en docencia (*student peer assessment*) nos referimos a actividades en las que unos alumnos emiten un juicio de valor sobre el trabajo realizado por algunos de sus compañeros, de acuerdo con criterios de calidad establecidos por el profesorado o por los propios alumnos [1].

Las actividades de evaluación entre compañeros tienen múltiples virtudes, entre las que podemos destacar las siguientes:

1. Al utilizar los criterios de calidad para evaluar otros trabajos, los alumnos los aprenden e interiorizan, de manera que los tendrán más presentes cuanto tengan que hacer un nuevo trabajo con los mismos criterios de calidad.
2. Los alumnos se esfuerzan más en hacer un trabajo de calidad si saben que éste va a ser evaluado por sus compañeros.
3. Los alumnos desarrollan la habilidad de emitir juicios de valor sobre el trabajo de otros, de acuerdo con unos criterios preestablecidos. Ésta es una competencia transversal muy útil en el entorno profesional.
4. Los alumnos se convierten en ayudantes del profesor para la dura tarea de evaluar las actividades del curso, bien sea con propósito formativo (determinar lo que hay que mejorar) o con propósito sumativo (acreditar ante terceros el nivel de aprendizaje alcanzado).
5. Los alumnos se esfuerzan en hacer un trabajo más inteligible, añadiendo explicaciones adicionales, consideradas hasta la fecha innecesarias pues el destinatario era un profesor "que todo lo sabe", pero sin duda imprescindibles cuando van a ser corregidos por compañeros cuya formación conocen mejor.

Estas virtudes se relacionan de forma muy directa con algunos de los requerimientos que se ponen sobre la mesa en el contexto del proyecto de construcción del Espacio Europeo de Educación Superior (EEES), especialmente en lo que respecta al desarrollo de nuevos modelos docentes centrados en el aprendizaje del alumno. En particular, se espera que, como consecuencia de la adaptación al EEES, nuestros planes de estudios presten más atención a determinadas habilidades transversales, como el desarrollo de un espíritu crítico frente a sus propios trabajos y los de los demás, señalada en el punto 3 de la lista anterior. Por otra parte, el modelo docente centrado en el aprendizaje requiere de la utilización de todos los medios a nuestro alcance para motivar al alumno a que realice la secuencia de actividades que le conducirán a los objetivos de aprendizaje (y que le ocuparán una significativa carga de horas, de acuerdo con la asignación de

créditos ECTS). Precisamente el punto número 2 de la lista anterior hace referencia a dicha motivación del alumno.

Finalmente y también en el contexto del modelo centrado en el aprendizaje, la evaluación continuada y especialmente los mecanismos de retroalimentación frecuente juegan un papel crucial. Teniendo en cuenta que estos mecanismos consumen mucho tiempo de dedicación por parte del profesor, toda la ayuda que podamos tener de los propios alumnos (tal y como se señala en el punto 4) es bienvenida.

No obstante, implicar a los alumnos en tareas de evaluación de compañeros de forma efectiva no es tarea fácil. Con frecuencia, actividades poco planificadas o directamente mal planteadas producen resultados decepcionantes (poco interés puesto de manifiesto por los alumnos, poca fiabilidad de las valoraciones, ligereza en la evaluación, etc.).

Los autores de este trabajo llevamos ya varios años utilizando la evaluación entre compañeros en el marco de dos asignaturas de enseñanza de la programación de ordenadores dentro del proceso de adaptación al EEES. En particular, los alumnos realizan una actividad para evaluar la versión final de un proyecto de programación realizado en equipo, de acuerdo con unos criterios especificados por el profesor al inicio del curso. Los resultados de esta evaluación son utilizados por el equipo de profesores para tomar decisiones de forma eficiente sobre la calificación final.

La experiencia acumulada nos ha ofrecido algunas claves que pueden marcar la diferencia entre una actividad que funcione y otra que no. El objetivo de este artículo es describir en detalle la forma en que se lleva a cabo la actividad de evaluación entre compañeros en nuestras asignaturas y subrayar esos aspectos que consideramos clave para su buen funcionamiento.

La sección 2 del artículo describe los aspectos generales de la organización de nuestras asignaturas. La sección 3 describe cómo son los proyectos a los que se aplica la actividad de evaluación entre compañeros, que se describe en detalle en la sección 4. La sección 5 muestra algunos resultados obtenidos en evaluaciones recientes, que nos dan algunas indicaciones sobre la fiabilidad y utilidad de la actividad. Finalmente, la sección 6 presenta las conclusiones de nuestro trabajo.

## **2. Organización de las asignaturas**

Aplicamos evaluación entre compañeros a dos asignaturas de programación de ordenadores que forman parte de los planes de estudios de las titulaciones de Ingeniería Técnica de telecomunicación e Ingeniería técnica aeronáutica, impartidas en la Escuela Politécnica Superior de Castelldefels (EPSC) [2], perteneciente a la Universidad Politécnica de Cataluña..

Las asignaturas se llaman *Introducción a los computadores (IC)* y *Laboratorio de programación (LP)*. Son comunes a los dos planes de estudios mencionados. La asignatura IC se imparte en el primer cuatrimestre del plan de estudios y se centra en los principios básicos de la programación en lenguaje C. La asignatura LP se imparte en el segundo cuatrimestre y completa lo aprendido en IC con una introducción a la programación orientada a objetos, programación visual, algoritmos y estructuras de datos más avanzados.

En virtud del proceso de adaptación de los planes de estudios al EEES [3] ambas asignaturas tienen 4,8 ECTS, lo cual implica que los alumnos deben dedicar en cada una de ellas una media de 8 horas de trabajo a la semana durante las 15 semanas que dura el

curso. De esas 8 horas semanales, 2 son de clase y 6 de trabajo fuera de clase, excepto 4 semanas en las que hay una sesión de 4 horas de clase (y, por tanto, sólo 4 horas de trabajo fuera de clase).

En ambas asignaturas los grupos de clase son de unos 40 alumnos que son atendidos por dos profesores. Esto representa una pequeña reducción en horas de clase para el profesorado, en relación al formato anterior de la asignatura en la que cada grupo de clase, también de 40 alumnos, requería la impartición, por parte de un profesor, de 6 horas de clase a la semana,

Los alumnos tienen un plan muy detallado de las actividades que deben realizar tanto en clase como fuera de ella, con un plan de entregas distribuidas a lo largo del curso, que permite al profesorado hacer un seguimiento del trabajo que realizan los alumnos y articular un proceso de retroalimentación frecuente y a tiempo.

En el plan de actividades de ambas asignaturas se hace un uso amplio de técnicas de aprendizaje cooperativo [4] (alumnos organizados en pequeños grupos para realizar juntos algunas de las tareas del programa) y técnicas de aprendizaje basado en proyectos [5] (los alumnos realizan en pequeños grupos un proyecto, aprendiendo por el camino todo lo que sea necesario). En particular, en ambas asignaturas los alumnos realizan un proyecto de programación en grupos de 3 o 4 personas. En IC el proyecto se desarrolla entre las semanas 9 y 15. En LP el proyecto es mucho más ambicioso y se desarrolla a lo largo de las 15 semanas del curso. Es justamente en el marco del desarrollo de estos proyectos en el que se realizan las actividades de evaluación entre compañeros en las que se centra este artículo. Más detalles sobre la organización de nuestras asignaturas, y más concretamente, sobre los criterios para su adaptación a los requerimientos del EEES pueden consultarse en [6].

### **3. Los proyectos y los criterios de calidad**

El proyecto que realizan los alumnos en IC consiste en una aplicación sencilla para trabajar con una lista de elementos básicos (por ejemplo, personas), y debe permitir realizar operaciones típicas de alta/baja/modificación de elementos de la lista. Con este proyecto los alumnos aprenden a trabajar con ficheros de texto, estructuras de datos simples y funciones.

El proyecto de LP es de mayor envergadura. Se trata de realizar una aplicación basada en objetos y con una interfaz visual atractiva, en la que deban utilizar varias estructuras de datos algo más complejas que las requeridas en IC (por ejemplo, pilas y colas), algoritmos de ordenación y gestión dinámica de memoria. A diferencia del proyecto de IC, en el que las funciones a implementar están completamente prescritas en el enunciado, en LP los grupos pueden incorporar al proyecto funciones adicionales dependiendo de sus intereses. En los últimos cursos, el proyecto de LP ha consistido en implementar una aplicación para la reserva de plazas en vuelos de pasajeros. En cursos anteriores el proyecto consistió en una aplicación para la resolución de Sudokus.

La forma en que se organiza el trabajo por proyectos es prácticamente igual en ambas asignaturas. Veamos a continuación las características más relevantes.

Los proyectos tienen unos criterios de calidad genéricos que se entregan a los alumnos en forma de rúbrica [7] en el momento en que se inicia el trabajo del proyecto. La figura 1 muestra la rúbrica para el proyecto de IC. En la columna de la izquierda se indican los 4 criterios de calidad (correcto, robusto, amigable y bien organizado y documentado) y

en las columnas de la derecha se indican tres niveles de calidad de referencia para cada criterio. La rúbrica indica incluso cómo va a puntuarse el resultado en función de los criterios de calidad.

La rúbrica para el proyecto de LP es muy similar, pero hay un criterio adicional asociado a la calidad de las funcionalidades extras que cada grupo añade por iniciativa propia al proyecto. Por lo tanto, en LP las calificaciones máximas asignadas a cada criterio son algo diferentes a las mostradas en la rúbrica de la figura 1.

Durante el proceso de realización de los proyectos, los grupos deben hacer una serie de entregas. Las dos más importantes son la demostración del primer prototipo y la versión final del proyecto.

El primer prototipo contiene ya los elementos clave del proyecto y permite al profesor detectar defectos importantes (que no haya detectado antes), de manera que los grupos aún tienen tiempo de reconducir el proyecto si fuera necesario. En IC el proyecto se inicia en la semana 9 y el primer prototipo debe entregarse en la sesión de clase de la semana 13. En LP el proyecto se inicia en la semana 1 y el primer prototipo debe estar listo en la semana 10. La evaluación del primer prototipo se realiza en base a los criterios establecidos en la rúbrica y la realiza el profesor, que además de establecer la calificación, emite un informe con comentarios para la mejora, sobre la base siempre de los criterios de calidad. Por tanto, se trata de una tarea de evaluación que requiere bastante tiempo para el profesor (unos 20 minutos por proyecto).

La versión final del proyecto debe entregarse en la sesión de la semana 15 (última del curso). Es precisamente en esa sesión de clase en la que los alumnos realizan la actividad de evaluación de los proyectos de algunos compañeros, que describiremos en detalle en la siguiente sección.

Finalmente, los proyectos tienen un peso del 40% en la calificación final de la asignatura. Esta calificación se determina de la siguiente forma:

- 10% correspondiente a la calificación del primer prototipo
- 20% correspondiente a la calificación de la versión final
- 10% correspondiente a una ampliación individual del proyecto

Todos los miembros del grupo reciben la misma calificación del primer prototipo y de la versión final. La ampliación individual es un ejercicio individual que realizan los alumnos en la última sesión de clase (semana 15). Se trata de una modificación del código del proyecto que resultará muy sencilla para aquellos alumnos que dominen la globalidad del proyecto (aunque se hayan especializado en una parte), resultará más difícil para los que se hayan concentrado en su parte del proyecto, sin prestar mucha atención al resto, y será imposible para los alumnos que ni siquiera se implicaron en su parte. La calificación de ese ejercicio de ampliación individual se determina de la siguiente forma:

- 0 si el alumno no hace bien la ampliación
- 5 si la hace bien
- 10 sólo en el caso de que todos los miembros del grupo hagan bien la ampliación individual

De esta forma se introduce un elemento de exigibilidad individual e interdependencia positiva que motiva a los alumnos a preocuparse de la globalidad del proyecto y a ayudar a sus compañeros de grupo [4].

El resto de los componentes de la nota final del curso son los siguientes:

- 20% por la realización a tiempo de todas las entregas del curso (si no se realizan al menos el 80% de las entregas, el curso está suspendido).
- 30% por la superación de los conocimientos básicos del curso, que se evalúan mediante exámenes individuales.
- 10% por actitud y participación, de acuerdo al criterio subjetivo de cada profesor (todas las asignaturas de la EPSC tienen esta componente en el método de calificación).

#### **4. La actividad de evaluación entre compañeros**

Tal y como se ha comentado ya, la actividad de evaluación entre compañeros se desarrolla durante la última sesión del curso (sesión 15), en la que cada grupo evalúa la calidad de la versión final del proyecto de otros dos grupos.

La sesión 15 es una de las cuatro sesiones de 4 horas de duración. El plan de esa sesión es el siguiente:

- 1 hora para que los grupos hagan los retoques finales al proyecto y una autoevaluación de acuerdo a los criterios de calidad.
- 1 hora para realizar la actividad de evaluación entre compañeros.
- Descanso de 15 minutos.
- 1 hora y 30 minutos para realizar el ejercicio de ampliación individual.
- 15 minutos para rellenar los cuestionarios finales de satisfacción (valoración del curso y del trabajo en grupo).

#### **Puesta a punto del proyecto**

Idealmente, los grupos llegan a la sesión 15 con el proyecto a punto, y en todo caso a falta de últimos retoques. Esto es así muchas veces, pero también ocurre con frecuencia que hay grupos que aún tienen partes del proyecto que no funcionan bien. En cualquier caso, la labor de los profesores durante la primera hora de la sesión es ayudar a los grupos a poner a punto los proyectos y resolver últimas dudas.

Al final de la primera hora de trabajo cada grupo debe decidir si está en condiciones de participar en la evaluación entre compañeros o no, valorando para ello en qué medida el proyecto satisface los criterios de calidad establecidos en la rúbrica. Esta decisión tiene diferentes consecuencias. Por una parte, si el grupo decide no participar en la actividad de evaluación entonces tendrá una hora más de trabajo para acabar de poner a punto el proyecto (mientras el resto de grupos hacen la actividad de evaluación). Sin embargo, el resultado de la evaluación entre compañeros es una entrega más del curso que no realizarán los grupos que no tengan el proyecto a punto. En algún caso, esta circunstancia podría poner en peligro el requerimiento de realizar como mínimo el 80% de todas las entregas. Por otra parte, el no tener el proyecto a tiempo para la evaluación

entre compañeros representa la pérdida de un punto en la calificación de la versión final del proyecto. Finalmente, si se participa en la evaluación entre compañeros con un proyecto que no está en condiciones entonces las valoraciones de los compañeros serán malas, lo cual afectará negativamente en el momento de la calificación final.

Este es un punto en el que la experiencia nos ha hecho cambiar cosas que no estábamos haciendo bien. En concreto, si bien la rúbrica que se ha mostrado en la figura 1 puede ser suficiente para orientar a los alumnos durante la realización del proyecto, no es suficientemente precisa para orientar a los alumnos en la decisión de participar o no en la evaluación entre compañeros. De hecho, la rúbrica tampoco es suficientemente precisa para orientar a los grupos en la tarea de evaluar los proyectos de los compañeros. Para mejorar la actividad, en las últimas ediciones, en la sesión 15 del curso usamos un documento mucho más preciso que denominamos *Criterios detallados para la evaluación de los proyectos*. El figura 2 muestra el ejemplo correspondiente a LP. El documento detalla el tipo de pruebas y criterios que deben usarse para determinar la calificación asignada cada uno de los criterios de calidad. De alguna manera, este documento completa la rúbrica con información que no es relevante durante la realización del proyecto pero que es crítica en el momento de su evaluación.

En resumen, el procedimiento mejorado que estamos utilizando en la actualidad es el siguiente:

- En la sesión 14 (penúltima del curso) distribuimos el documento de *Criterios detallados para la evaluación del proyecto*, de manera que los grupos pueden acabar de poner a punto su proyecto durante la semana, teniendo en cuenta esos criterios.
- Durante la primera hora de la sesión 15 los grupos acaban de poner a punto su proyecto y realizan una autoevaluación aplicando los criterios detallados de evaluación distribuidos la semana anterior. En esta autoevaluación simplemente tiene que rellenar un informe en el que indicarán la calificación que se asignan a cada criterio y la justificación correspondiente, usando siempre como referencia los criterios detallados de evaluación.
- Los grupos toman la decisión de participar o no en la evaluación entre compañeros, sobre la base del resultado de su autoevaluación. Podríamos incluso establecer como criterio que para participar en la evaluación deben haber obtenido una calificación mínima determinada en la autoevaluación (aunque de momento no hemos fijado esa calificación mínima y dejamos a los grupos que tomen la decisión). Los grupos que deciden participar en la evaluación deben enviar al campus digital una copia de la versión final del proyecto al finalizar la primera hora de clase de la sesión 15.

### **La evaluación cruzada**

En términos generales, la evaluación cruzada puede tener asociada dos tipos de calificaciones:

1. la calificación que otorgan los estudiantes a un determinado ejercicio, en este caso el proyecto del curso;
2. la calificación asociada al propio ejercicio de evaluar;

En el primer punto, el hecho de evaluar a los demás (e incluso a uno mismo) ayuda a los alumnos a centrar la atención en los criterios aplicados a la hora de poner la nota. Necesariamente, ello repercutirá en un esfuerzo mayor sobre los objetivos clave de la asignatura para aquellos alumnos interesados en mejorar su calificación. A modo de ejemplo, los estudiantes suelen tener tendencia a prestar mayor atención al aspecto estético del programa (la amigabilidad) que a la documentación, por lo que los programas suelen estar muy trabajados en la manera de presentar sus menús y opciones para el usuario pero muy poco, por no decir nada, documentados. Si en los criterios se indicara un peso de 5 puntos sobre 10 a la documentación y de 1 punto sobre 10 a la amigabilidad, puede modularse esta tendencia natural, efecto que se acentúa cuando los alumnos deben asignar una nota muy baja a compañeros que han presentado un programa espectacular en su programación visual pero sin ningún comentario.

En el segundo punto, existe el riesgo de que los alumnos decidan emplear otros criterios a los señalados para valorar el trabajo de sus compañeros, especialmente si desean favorecer a sus amistades, penalizar a sus contrarios o, sencillamente, no dedicar esfuerzo a tal ejercicio. Definir la manera adecuada de corregir no es sencillo, salvo el caso en que tratemos de ejercicios tipo test, pero una regla que suele ser suficiente para los alumnos es indicar que si su valoración difiere significativamente de la del profesor serán penalizados, mientras que si es similar serán recompensados, todo ello según la rúbrica mencionada, por supuesto.

En nuestro caso, la manera en que implantamos este proceso es la siguiente:

Después de la primera hora, y una vez se tiene la lista de grupos que van a participar en la evaluación entre compañeros, empieza propiamente la actividad. Cada grupo evaluará a otros dos grupos, dedicando 30 minutos a cada evaluación. El profesor escribe en la pizarra una tabla que indica cómo se han asignado las evaluaciones (por ejemplo, el grupo 1 evaluará a los grupos 2 y 3, el grupo 2 a los grupos 3 y 4, etc.). Para facilitar el proceso, cada grupo que participa en la evaluación marcará claramente el ordenador en el que está disponible su proyecto (cada grupo trabaja normalmente con dos o tres ordenadores portátiles). Para ello usa un folio que indica además su número de grupo. Además, puesto que durante el proceso de evaluación los códigos y ficheros de datos pueden sufrir alteraciones, cada grupo debe tener preparada una copia de seguridad de su proyecto por si fuera necesario volver a ponerlo a punto. De hecho, después de la primera evaluación, cada grupo podrá regresar unos minutos al ordenador en el que está su proyecto para dejarlo todo listo de cara a la segunda evaluación.

Antes de empezar la primera evaluación, el profesor reparte a cada grupo cuatro copias del formulario de evaluación. En ese formulario (similar al de autoevaluación) los grupos deben indicar el número del grupo evaluador, el número del grupo evaluado, la calificación asignada a cada criterio, con una breve justificación de las razones de esa calificación, y la calificación final (ver figura 3). Para cada evaluación los grupos han de producir dos copias del resultado (por eso se necesitan cuatro copias del formulario). Una de las copias de cada evaluación será para el profesor y la otra para el grupo evaluado, aunque esta segunda copia de la evaluación se entregará al grupo evaluado al final de todo el proceso. Puesto que cada grupo es de tres o cuatro alumnos se les recomienda que, a medida que van tomando decisiones, dos miembros del grupo vayan rellenando en paralelo las dos copias del formulario de evaluación.

Es importante advertir a los alumnos que, si bien la calificación de la versión final del proyecto será decisión del profesor, los resultados de la evaluación entre compañeros se tendrán en cuenta. Además, el interés en realizar una evaluación correcta se premiará a

través de la componente de actitud y participación, que, como hemos visto antes, tiene un peso del 10 % en la calificación de la asignatura.

Con todas estas instrucciones y consideraciones se desarrolla la actividad durante una hora (30 minutos para cada evaluación). Durante la actividad conviene evitar que los grupos interactúen entre sí excesivamente, porque cada grupo debería evaluar lo que ve, sin necesidad de explicaciones de los autores. El profesor recoge las dos hojas por grupo después de cada evaluación y al final del proceso entrega a cada grupo una copia de las dos evaluaciones realizadas a su proyecto.

La sesión continúa con un descanso, la prueba de ampliación individual y los cuestionarios finales, que son elementos ya ajenos a la actividad de evaluación que estamos considerando en este artículo.

### **La calificación del proyecto**

Con todo este proceso, los profesores obtenemos, para cada proyecto, una nota de autoevaluación y dos notas de evaluaciones de otros grupos. Si esas tres calificaciones coinciden razonablemente entre sí y coinciden también con la percepción que tiene el profesor en relación a la calidad del proyecto (que conoce porque lo ha estado supervisando durante gran parte del curso) entonces se asigna al proyecto la nota media de las evaluaciones (recordemos que esa calificación tiene un peso del 20% en la nota final de la asignatura). En el caso de que existan discrepancias significativas entre las calificaciones (entre sí y/o con la percepción del profesor) entonces el profesor evalúa el proyecto personalmente y decide la calificación final (aunque las evaluaciones sean discrepantes, las observaciones reflejadas por los alumnos en sus evaluaciones normalmente ayudan al profesor a ser más ágil en su evaluación). Si a partir de su evaluación determina que algún grupo no se ha tomado en serio la autoevaluación o la evaluación de los compañeros, entonces toma las decisiones oportunas en relación a la calificación de actitud y participación.

## **5. Algunos datos**

La figura 4 muestra datos procedentes de la evaluación de los proyectos de 22 grupos (de 3 o 4 alumnos cada grupo) correspondientes a 2 grupos de clase de la asignatura LP en el segundo cuatrimestre del curso 2007-2008. Se trata de todos los proyectos de ese curso en esos dos grupos de clase excepto uno que no participó en la evaluación entre compañeros por decisión de sus autores.

Los proyectos están numerados del 1 al 22 en las columnas G. Para cada proyecto se aportan tres filas de calificaciones correspondientes a la autoevaluación (la fila de números en *cursiva*) y las dos evaluaciones realizadas por otros grupos. Las columnas CO, FA, RO, AM y OD indican las calificaciones asignadas a cada uno de los criterios de calidad: Correcto, Funcionalidades adicionales, Robusto, Amigable, Organizado y Documentado. Las columnas T indican la nota total (suma de las notas de los diferentes criterios). En la cabecera de la tabla se indica la calificación máxima que puede obtenerse en los diferentes criterios y en la nota total. Finalmente, las columnas P indican la calificación asignada por el profesor al proyecto.

G	CO 3	FA 2	RO 1,5	AM 1,5	OD 2	T 10	P	G	CO 3	FA 2	RO 1,5	AM 1,5	OD 2	T 10	P
<b>1</b>	1	0	1,5	1,25	1,25	5	<b>6,5</b>	<b>12</b>	2	1	1,25	1,5	1	6,75	<b>8,5</b>
	1,5	1	0,5	1,5	1,5	6			2	1,5	1	0,75	1	6,25	
	2	1,5	1,5	1,5	1	7,5			2,5	2	1,25	1	0,5	7,25	
<b>2</b>	2,5	1,5	1,5	1,5	2	9	<b>9</b>	<b>13</b>	3	1,5	1,5	1,5	2	9,5	<b>10</b>
	2	1,5	1,5	1,5	2	8,5			3	2	1,5	1	2	9,5	
	2	2	1	1,5	1,5	8			3	2	1,5	1,5	2	10	
<b>3</b>	2	1	1,5	1	1,5	7	<b>7</b>	<b>14</b>	1,5	1	0,5	0,75	1,25	5	<b>5</b>
	2	1	0,75	1	1	5,75			1,5	0,25	0,25	1,5	1,5	5	
	1,5	0,5	0,5	1	1	4,5			2	0,5	0,75	1	0,75	5	
<b>4</b>	2	2	0,5	1,5	1	7	<b>7</b>	<b>15</b>	2,75	1,25	1,5	1,5	2	9	<b>8,5</b>
	2,5	1	0,75	1,5	1,5	7,25			3	0,5	1,5	1,5	0,75	7,25	
	1	2	0,5	0,75	2	6,25			3	1	1,5	1,5	1,5	8,5	
<b>5</b>	2	1	1,5	1,5	1,5	7,5	<b>7,5</b>	<b>16</b>	2,5	1,5	0,5	1	1,5	7	<b>7</b>
	2,5	1	1,5	1	1,5	7,5			2,25	0,25	0,25	1	1,25	5	
	2	1	1,5	1	2	7,5			3	0,5	1	1	1	6,5	
<b>6</b>	3	2	1,5	1,5	2	10	<b>9,5</b>	<b>17</b>	2	1	1,5	1,25	1,5	7,25	<b>7</b>
	2,5	2	1,5	1	2	9			2	0,5	1	1,25	1,5	6,25	
	3	2	1,5	1,5	2	10			3	1,5	1,5	1,5	0,5	8	
<b>7</b>	3	1,5	1,5	1	1,25	8,25	<b>7</b>	<b>18</b>	1,5	0	1,5	1,5	1	5,5	<b>5,5</b>
	1,5	2	1	0,75	1	6,25			2	0,5	1	1,5	1,5	6,5	
	2	2	1,5	0,75	1,5	7,75			1,5	1	1,5	1	1	6	
<b>8</b>	2	0,5	1,25	1,5	1,75	7	<b>5,5</b>	<b>19</b>	1	0,75	1,5	1,5	1,5	6,25	<b>5</b>
	2	1	0,5	1,5	1	6			1	1,5	1	1	2	6,5	
	1	1	0,5	1,25	2	5,75			2	1	0	1,5	1	5,5	
<b>9</b>	0,5	0	0,5	1,5	1	3,5	<b>5</b>	<b>20</b>	2,5	0,75	1	1,5	1,75	7,5	<b>9</b>
	0,5	0,5	1	1	2	5			2,5	2	1,5	1	0,5	7,5	
	1	1	1	1	2	6			2,5	0,5	0,75	1	1,75	6,5	
<b>10</b>	1,5	1	1,5	0,75	2	6,75	<b>4,5</b>	<b>21</b>	2	1	1,5	1	1,5	7	<b>6,5</b>
	2	0,5	1,25	1,5	1,5	6,75			2	1,25	1	1,25	2	7,5	
	1	0	0,5	1	1,5	4			2	1,75	1	1	2	7,75	
<b>11</b>	3	2	1,3	1,3	1,8	9,5	<b>9</b>	<b>22</b>	1	0,5	1	1,5	1	5	<b>5,5</b>
	2,75	2	1	1,5	1,5	8,75			3	1,5	1	1	2	8,5	
	2	2	1,5	1,5	2	9			1	1,25	0,25	1,5	1,5	5,5	

**Figura 4:** Datos de las evaluaciones cruzadas correspondientes al proyecto de LP durante al curso 2007-2008

Una primera observación interesante de estos datos es que todos los proyectos presentados alcanzan una nota de aprobado o superior. Esto es un resultado normal cuando se trabaja por proyectos y existe un seguimiento adecuado por parte del profesorado. En otras palabras, durante el curso se toman las medidas oportunas para que el proyecto acabe bien. Esto no quiere decir que todos los alumnos aprueben la asignatura. Hay alumnos que abandonan la asignatura por el camino, porque no disponen del tiempo de dedicación exigido y algunos alumnos suspenden la asignatura (aunque hayan aprobado el proyecto) porque no alcanzan los conocimientos básicos establecidos, tal y como se comentó en la sección 3.

Veamos ahora algunas observaciones que tienen que ver con el nivel de discrepancia entre la autoevaluación, evaluaciones de compañeros y valoración subjetiva del profesor:

1. En un total de 8 casos (proyectos 2, 4, 5, 6, 11, 13, 14 y 18) existe coincidencia entre la autoevaluación, las dos evaluaciones de compañeros (diferencias de un punto o menos entre las calificaciones totales de esas tres evaluaciones) y la valoración subjetiva del profesor. Estos son los casos en los que el profesor requiere poco tiempo para decidir la calificación que asigna al proyecto.
2. Hay 4 casos (proyectos 12, 19, 20 y 21) en los que hay coincidencia entre la autoevaluación y las dos evaluaciones de compañeros, pero no con la valoración

subjetiva del profesor, que tuvo que dedicar un tiempo a decidir la nota final. En dos de estos casos (proyectos 12 y 20) la nota del profesor fue superior y en los otros dos fue inferior.

3. En 6 casos (proyectos 1, 3, 8, 9, 10 y 22) hay gran discrepancia entre la autoevaluación y las dos evaluaciones de compañeros (diferencias de más de dos puntos entre las notas totales).
4. En 4 casos (proyectos 1, 3, 8 y 9) la calificación discrepante es la de la autoevaluación. En el caso de los proyectos 1 y 9 la autoevaluación es claramente inferior y no coincide con la valoración del profesor (grupos muy autoexigentes). En el caso de proyecto 8 la autoevaluación está claramente por encima y no coincide con el profesor (grupo autocomplaciente). En el caso del proyecto 3 la autoevaluación está claramente por encima y coincide con el profesor (un grupo injustamente tratado por los compañeros).
5. Hay 2 casos en los que la discrepancia es debida a una de las dos evaluaciones de compañeros. En el caso del proyecto 10 la evaluación discrepante coincide con la del profesor. En el caso del proyecto 22 la evaluación discrepante no coincide con la del profesor.
6. Hay 5 casos (proyectos 7, 15, 16 y 17) en los que hay discrepancias pero no son debidas a una nota que se diferencia mucho de las otras dos.
7. En todos los casos excepto en 5 (proyectos 3, 10, 12, 19 y 20) la diferencia entre calificación del profesor y la nota que se obtendría como media entre autoevaluación y evaluación entre compañeros es menor o igual a 1 punto (nótese que ello supone 0,2 puntos sobre el total de la nota final de curso).

Finalmente, nos interesa analizar el grado de discrepancia en las valoraciones de los diferentes criterios de calidad. Para ello, asociamos a cada criterio un valor que se obtiene calculando, para cada proyecto, la diferencia máxima entre las tres calificaciones asignadas a ese criterio, y calculando después la media de esas diferencias máximas dividida por la calificación máxima posible para el criterio en cuestión.

Los valores obtenidos son:

*Correcto: 0,23*

*Funciones adicionales: 0,37*

*Robusto: 0,36*

*Amigable: 0,23*

*Organizado y bien documentado: 0,38*

Idealmente estos valores deberían ser 0, lo cual indicaría que siempre hay coincidencia entre todas las calificaciones. En el peor caso los valores serían 1, indicando que en cada caso siempre hay una calificación que toma el valor máximo posible y otra que es 0. Veamos ahora que nos dicen los resultados obtenidos.

En el caso del criterio *Correcto* las pruebas a realizar durante la evaluación están descritas de forma exhaustiva en el documento de criterios detallados de evaluación. Normalmente, los alumnos harán esas pruebas y no otras durante la evaluación. Es razonable que sea uno de los criterios en que más coincidencia hay.

El criterio *Funcionalidades adicionales* es uno de los más subjetivos, porque se valora el grado de atractivo de las funcionalidades de libre elección de cada grupo, a menudo

por encima de la complejidad técnica, más valorada por los profesores. Es lógico, por tanto, que sea uno de los criterios que más discrepancias presenta.

En el caso del criterio *Robusto*, las pruebas que hay que realizar durante la evaluación también están explicitadas pero no son muy exhaustivas. Es razonable suponer que durante las pruebas de evaluación ocurren más errores de los identificados en la lista. Como el criterio está definido en relación a estos dos posibles errores, es probable que los alumnos no tengan muy claro cómo calificar si hay otros errores que no aparecen en la lista.

El criterio *Amigable* es relativamente subjetivo, por lo dificultoso de su definición, pero no en términos comparativos, donde los estudiantes intuyen con facilidad las aplicaciones que les resultan más amigables.

Finalmente, en el criterio *Organizado y bien documentado* se produce el mayor nivel de discrepancia. Probablemente, la primera parte de las instrucciones para la evaluación de este criterio (ver figura 3) es suficientemente objetiva. Sin embargo, no lo es la segunda parte, que hace referencia a la claridad de los comentarios. Por un lado, el concepto de claridad es algo subjetivo (un comentario puede ser claro para unos y no para otros). Pero además es posible que unos grupos miren una parte del código y otros miren otra parte diferente y su percepción sobre la claridad de los comentarios sea diferente. Está claro que las instrucciones para evaluar este criterio pueden mejorarse, indicando, por ejemplo, en qué partes del código debe fijarse el grupo evaluador para valorar la claridad de los comentarios.

## 6. Conclusiones

En este trabajo hemos descrito en detalle cómo llevamos a cabo una actividad de evaluación entre compañeros, en el contexto del desarrollo de un proyecto de programación de computadores, haciendo énfasis en los elementos que nos parecen más importantes para que la actividad sea productiva.

De acuerdo con nuestra experiencia, una cuestión clave es la definición de los criterios de evaluación. Es básico que esos criterios sean claros y objetivos. Inicialmente usábamos rúbricas como la mostrada en la figura 1, pero rápidamente vimos que, si bien la rúbrica juega muy bien su papel para proyectar unos criterios de calidad generales desde el inicio del proyecto, no es suficientemente detallada como guía para la evaluación, con lo que la fiabilidad de las evaluaciones era limitada. Actualmente usamos guías detalladas de evaluación, como la mostrada en la figura 2. Estas guías deben ser revisadas y mejoradas frecuentemente para aumentar la fiabilidad del proceso. Así por ejemplo, en la sección 5 hemos visto cómo los datos nos señalan la necesidad de mejorar las instrucciones para la evaluación de los criterios *Robusto* y *Documentado* y *bien organizado*.

Otro aspecto importante es recoger varias evaluaciones de cada proyecto (en nuestro caso, una autoevaluación y dos evaluaciones entre compañeros). De esta forma, una mera comparación entre las diferentes evaluaciones de un mismo proyecto puede ayudarnos a valorar rápidamente la fiabilidad del proceso. En las primeras ocasiones, cada proyecto era evaluado por un solo grupo, y con una sola nota es más difícil determinar si el resultado es fiable o no.

Del análisis realizado en el apartado 5 correspondiente a un caso concreto nos interesa resaltar que en más de una tercera parte de los casos hay coincidencia entre todas las

valoraciones (incluida la subjetiva del profesor). En esos casos se agiliza mucho la decisión de la calificación final. Es interesante también la observación 7, que nos indica que si el profesor hubiese puesto como nota del proyecto la media entre las tres calificaciones asignadas por los alumnos, la nota final apenas habría variado. Ello simplifica las tareas de evaluación que terminan requiriendo poco esfuerzo por su parte, de forma que el esfuerzo ahorrado puede invertirse en otras muchas tareas del curso.

Conviene recordar que, tal y como se indicó en la introducción, las actividades de evaluación entre compañeros no sólo nos ayudan a tomar decisiones sobre la calificación del trabajo, sino que introducen elementos de motivación facilitan la generación de retroalimentación frecuente y a tiempo, ayudan a los alumnos a interiorizar los criterios y estimulan, por tanto, su pensamiento y capacidad crítica. En nuestro ejemplo, los criterios de calidad de los proyectos de LP son muy similares a los criterios de los proyectos de IC (la asignatura anterior), con lo que la evaluación que realizan los alumnos al final de IC les ayuda a comprender mejor esos criterios de cara a la realización del proyecto en LP. No obstante, hay que reconocer que, tal y como estamos haciendo ahora la actividad, su valor formativo podría mejorarse, ya que la evaluación se hace el último día de clase, con lo que los alumnos no tienen oportunidad de mejorar su proyecto a partir de lo que han aprendido durante la evaluación. Para reforzar su valor formativo, la evaluación entre compañeros debería hacerse también con versiones anteriores del proyecto. Una primera intuición del valor de esta actividad más elaborada es la evolución apreciada entre los resultados del ejercicio realizado durante el primer cuatrimestre, en el marco de IC, y los obtenidos durante el segundo cuatrimestre, en el marco de LP, con justificaciones a las evaluaciones realizadas por los alumnos mucho más precisas y atinadas.

Finalmente, conviene indicar que actividades de evaluación entre compañeros usando procedimientos similares a los descritos aquí pueden realizarse con otros tipos de productos del trabajo de los alumnos, como es el caso de informes escritos o presentaciones orales. Los autores tienen algunas experiencias positivas al respecto, aunque en el marco de otras asignaturas del plan de estudios.

## Referencias

- [1] A.W. Bangert, "Peer Assessment: A Win-Win Instructional Strategy for Both Students and Teachers," *J. Cooperation & Collaboration in College Teaching*, Vol. 10, No. 2, p. 77.
- [2] Escola Politècnica Superior de Castelldefels (EPSC), <http://epsc.upc.edu/ca/>
- [3] L. Alonso , C. Collado, J. Hernandez, S. Ruiz, M. Valero-García, "La prueba piloto de adaptación al EEES en la Escuela Politècnica Superior de Castelldefels", *Jornada Nacional sobre Estudios Universitarios*, Castellón 2008
- [4] D. W. Johnson, R. T. Johnson y K. A. Smith. *Active learning: Cooperative Learning in the college classroom*. Edina MN: Interaction Book Company, 1991.
- [5] D.R. Woods et al. "The future of engineering education. Developing critical skills". *Chem Eng. Education*, 34(2), 108-117

- [6] J. Anguas et al, “Una experiencia de adaptación al EEES de dos asignaturas de programación de ordenadores”, *Cuarto Congreso Internacional Docencia Universitaria e Innovación (CIDUI 2006)*.
- [7] Ideas and Rubric, Instructional Intranet, Chicago Public Schools, [http://intranet.cps.k12.il.us/Assessments/Ideas\\_and\\_Rubrics/ideas\\_and\\_rubrics.html](http://intranet.cps.k12.il.us/Assessments/Ideas_and_Rubrics/ideas_and_rubrics.html)

<b>Criterios para la evaluación de los prototipos</b>			
	<b>Nivel de calidad</b>		
<b>Criterio</b>	<b>Bien</b>	<b>Mejorable</b>	<b>Mal (0 puntos)</b>
<b>Correcto (4 puntos)</b>	La aplicación realiza correctamente todas las funciones previstas en el enunciado, y ha funcionado bien con todas las pruebas que he hecho (ha fallado como máximo una vez), y eso que he hecho muchas pruebas <b>(4 puntos)</b> .	Falta una de las funciones previstas, o bien la aplicación ha fallado en tres o cuatro ocasiones <b>(2 puntos)</b> .	Faltan dos o más de las funciones previstas, o bien la aplicación falla con mucha frecuencia. Está claro que no está bien.
<b>Robusto (1,5 puntos)</b>	La aplicación resiste sin bloquearse todos los errores típicos que pueden aparecer. No he conseguido que se cuelgue <b>(1,5 puntos)</b> .	Es razonablemente robusto. No es fácil que se quede colgado, pero en uno o dos casos se bloqueó <b>(0,75 puntos)</b> .	La aplicación no es robusta en absoluto. Se queda colgada con frecuencia ante errores típicos.
<b>Amigable (1,5 puntos)</b>	El usuario no tiene ninguna duda, en ningún momento, sobre cómo interactuar con la aplicación, qué datos debe entrar y cómo, y cómo interpretar los resultados y mensajes de la aplicación <b>(1,5 puntos)</b> .	Los mensajes e información que da la aplicación son suficientes para trabajar bien. Sin embargo, en alguna ocasión he tenido algunas dudas sobre lo que hay que hacer o cómo hay que hacerlo <b>(0,75 puntos)</b> .	El usuario tiene dudas constantes sobre lo que le está pidiendo la aplicación, y es difícil interpretar los resultados y mensajes en pantalla.
<b>Bien organizado y documentado (3 puntos)</b>	El código está bien organizado. Es muy fácil encontrar el punto de la aplicación que hay que tocar para realizar alguna modificación en la funcionalidad. Cada procedimiento y función tiene un comentario inicial que explica lo que hace, y cuáles son los parámetros. Además, los puntos del código especialmente complicados tienen un comentario suficientemente clarificador. Las variables, procedimientos y funciones tienen nombres que ayudan a comprender para qué se usan. El código está bien indentado <b>(3 puntos)</b> .	Está razonablemente bien organizado y documentado, aunque en algún caso, la estructuración en procedimientos y funciones podría ser mejor. Los comentarios son suficientes, aunque echo de menos alguna aclaración más en algún punto del código <b>(1,5 puntos)</b> .	La estructura del código no tiene lógica, y no hay comentarios (o los que hay no clarifican nada). El código no está bien indentado. Sería incapaz de modificar este código para añadir alguna funcionalidad nueva o arreglar algún error.

Figura 1: Rúbrica de criterios de calidad para el proyecto de IC.

### Detalles para la evaluación del proyecto: Air Madrid: Ahora si.

**Previo:** Es importante que cada proyecto **tenga una opción** llamada Acerca de (o similar) en la que se muestre los **nombres de los autores** (y si es posible su foto) y se indiquen claramente cuáles son las **funcionalidades extras** que el grupo ha incorporado al proyecto).

#### Normas para la evaluación

Para evaluar el criterio **Correcto**, verificar que la aplicación puede realizar las operaciones siguientes:

El cliente:	Seleccionar un vuelo Reservar dos plazas Deshacer la segunda reserva realizada
El encargado:	Seleccionar un vuelo (el mismo que seleccionó el cliente) Comprobar los datos reserva que realizó el cliente Anular la reserva Hacer una nueva reserva Obtener una lista de pasajeros del vuelo Ordenar la lista de pasajeros Crear un vuelo nuevo vacío
De nuevo el cliente:	Seleccionar el vuelo que el encargado acaba de dar de alta Reservar una plaza en ese vuelo

Además, comprobar que al menos la pila necesaria para la operación de deshacer utiliza una gestión dinámica de la memoria.

Asignar **3 puntos** a este criterio sólo si la aplicación realiza de forma correcta todas estas operaciones. Asignar **2 puntos** si una de esas operaciones no funciona bien. Asignar **1 punto** si dos o tres de las operaciones no funcionan y asignar **0 puntos** si fallan más de tres operaciones.

Para evaluar el criterio **Funcionalidades adicionales** utilizar la tabla de criterios de calidad:

**2 puntos:** La aplicación tiene muchas funcionalidades adicionales interesantes, que funcionan correctamente. Si yo fuese el director de la empresa, cogería esta aplicación (casi no necesito ver más para elegir).

**1 punto:** La aplicación no está mal. Tiene alguna funcionalidad que está bien, pero creo que podría mejorarse bastante. Si fuese director de la empresa, me gustaría ver más posibilidades.

**0 puntos:** La aplicación es pobre. Se ha limitado a las funciones mínimas requeridas. Se nota que para este grupo el proyecto no era prioritario.

Para evaluar el criterio **Robusto**, verificar al menos que la aplicación resiste los siguientes errores:

- Formato incorrecto en los datos del fichero de entrada (por ejemplo, falta algún dato)
- El usuario ha introducido una palabra cuando debía haber escrito su edad

Asignar **1,5 puntos** si resiste bien todos los errores. Asignar **0,5 puntos** si falla en uno de los dos casos anteriores. Asignar **0 puntos** si falla en ambos casos.

Para evaluar el criterio **Amigable** utilizar la tabla de criterios de calidad:

**1,5 puntos:** El usuario no tiene ninguna duda, en ningún momento, sobre cómo interactuar con la aplicación, qué datos debe entrar y cómo, y cómo interpretar los resultados y mensajes de la aplicación

**1 punto:** Los mensajes e información que da la aplicación son suficientes para trabajar bien. Sin embargo, alguna vez he tenido dudas sobre lo que hay que hacer o cómo hay que hacerlo.

**0 puntos:** El usuario tiene dudas constantes sobre lo que le está pidiendo la aplicación, y es difícil interpretar los resultados y mensajes en pantalla.

Para evaluar el criterio **Bien organizado y documentado**

- Asignar **un punto** a la valoración de la organización del código. Para ello, comprobar por ejemplo, que es muy fácil localizar:
  - el método donde se ordenan los pasajeros de un vuelo,
  - el sitio en que el encargado da de alta un nuevo vuelo
  - el sitio donde se comprueba que datos introducidos por el usuario no contienen errores.
- Asignar **un punto** a los comentarios (existen, son claros y pertinentes) e indentación, tanto en el código de Visual Basic como en el de Visual C++

**Figura 2:** Criterios detallados para la evaluación del proyecto.

<b>Plantilla para la evaluación del proyecto</b>	
<b>Nombre de los miembros del grupo</b>	
<b>Instrucciones</b>	
Calificad el proyecto asignado, indicando vuestra puntuación para cada uno de los criterios de calidad. Justificad brevemente vuestras valoraciones.	
Debéis hacer dos copias de cada valoración, una para el profesor y otra para el grupo evaluado.	
Tened en cuenta que la calidad de vuestra evaluación será tenida en cuenta en la valoración subjetiva que realizará el profesor.	
<b>Grupo evaluado #</b>	
<b>Correcto (3)</b>	¿por qué?
Calificación	
<b>Funciones adicionales (2)</b>	¿por qué?
Calificación	
<b>Robusto (1,5)</b>	¿por qué?
Calificación	
<b>Amigable (1,5)</b>	¿por qué?
Calificación	
<b>Bien organizado y documentado (2)</b>	¿por qué?
Calificación	
<b>CALIFICACIÓN TOTAL</b>	Este proyecto, comparado con el vuestro es: ¿Mejor, similar, peor? ¿Por qué?

**Figura 3:** Plantilla para la evaluación cruzada (apta también para la autoevaluación).