

Una metodología para fomentar el aprendizaje mediante sistemas de evaluación entre pares

Pablo Sánchez

Carlos Blanco

Dpto. Matemáticas, Estadística y Computación

Universidad de Cantabria

Santander(Cantabria)

p.sanchez@unican.es

carlos.blanco@unican.es

Resumen

Como consecuencia de la implantación de las directrices del *Espacio Europeo de Educación Superior* se han adoptado en diferentes cursos universitarios diversos sistemas de evaluación continua. Uno de los principales problemas asociados a dichos sistemas es el incremento en la carga del trabajo del docente, la cual puede llegar a ser insostenible. Con el objetivo de aliviar dicha carga de trabajo, se han propuesto diferentes estrategias de evaluación entre iguales. Se pretende además que dichas estrategias contribuyan de forma efectiva al aprendizaje. Para ello, tales estrategias deben de estar convenientemente diseñadas. Además, debemos asegurarnos de que al transferir la actividad evaluadora del docente al alumnado, no estamos sobrecargando a este último. Para solventar estos problemas, este artículo presenta una metodología concreta que permite organizar sistemas de evaluación entre iguales de manera que: (1) se asegure que la actividad evaluadora se realiza con rigor; (2) se fomente el aprendizaje; y (3) se asigne una carga de trabajo justa y adecuada a cada alumno. Dicha metodología se ha aplicado con éxito a una asignatura de *Ingeniería del Software* durante los dos últimos cursos académicos.

Abstract

In last years, many universities have adopted, according to the guideline of the *European Space of Higher Education*, different strategies for continuous assessment. A noticeable shortcoming of these strategies is the increase in instructors' workload, which can become unfeasible. To improve this situation, different *peer assessment* techniques have been designed. It is expected that these strategies also promote an effective learning. However, to reach this goal, they must be appropriately designed. In addition, it must be ensured that the assessment quality does not decrease as a consequence of being carried out by students instead

of instructors. Finally, it must be also checked that students' workload also remains in a reasonable frame. This article presents a methodology for preparing *peer assessment activities* so that: (1) effectiveness of produced assessments can be assured; (2) they promote *effective learning*; (3) ensures a fair workload to each student, avoiding overloading. This methodology has been successfully applied to a *Software Engineering* subject during the two last academic years.

Palabras clave

Evaluación entre Iguales, Aprendizaje Basado en Proyectos, Ingeniería del Software.

1. Introducción

Las nuevas directrices del *Espacio Europeo de Educación Superior* han hecho que cada día más instructores universitarios adopten técnicas de aprendizaje activo [1] y evaluación continua [10]. De acuerdo con dichas técnicas, los instructores deben diseñar y proponer una serie de actividades enfocadas a fomentar y propiciar el aprendizaje del alumno. Lo normal es que el alumno yerre durante el desarrollo de dichas actividades. Por tanto, si no se supervisan las actividades de los alumnos y se corrigen sus fallos, muy difícilmente se producirá el aprendizaje, pues los alumnos seguirán errando una y otra vez. Por tanto, para que se produzca un verdadero aprendizaje efectivo, los equipos docentes deben monitorizar y evaluar, con carácter formativo, los resultados de las diferentes actividades propuestas a los alumnos. Este aumento en la carga de trabajo del equipo docente puede llevar a procesos de enseñanza-aprendizaje inabordables que generen situaciones de estrés y ansiedad indeseadas, tanto en los equipos docentes como del alumnado [7].

Una de las soluciones que se han venido proporcionando en los últimos años para solventar este problema es trasladar parte del trabajo de evaluación del pro-

Req01	El alumnado obtiene una evaluación formativa similar a la que obtendría del profesorado
Req02	La actividad evaluadora no sobrecarga en exceso al alumnado
Req03	La actividad evaluadora está diseñada para que fomente el aprendizaje

Cuadro 1: Requisitos para una adecuada *evaluación por pares*

fesorado al alumnado. El alumnado, asistido por *rúbricas* [9] u otros mecanismos, se autoevalúa o evalúa a sus compañeros. Este último esquema es el que se conoce como *evaluación entre iguales o por pares* [4, 5, 10].

Los promotores de este tipo de evaluación argumentan que se trata además de una actividad de aprendizaje. El alumno, mediante la evaluación de su trabajo o el de sus compañeros, identifica sus propios errores y aprende a solucionarlos. No obstante, la evaluación entre pares plantea una serie de problemas que deben ser tenidos en cuenta y solventados si queremos conseguir estos beneficios.

En opinión de los autores de este documento, para que un sistema de evaluación entre pares sea efectivo y contribuya al aprendizaje, éste debe satisfacer los tres requisitos que se exponen en el Cuadro 1.

El requisito *Req-01* trata de asegurar que el traslado la actividad evaluadora del profesorado al alumnado no vaya en perjuicio del propio alumnado. Para ello tenemos que corroborar que el alumnado está capacitado para realizar la actividad evaluadora. Además, debemos verificar que realiza dicha actividad con el rigor y la seriedad que merece.

En el caso del requisito *Req-02*, se trata de evitar situaciones indeseadas de estrés y ansiedad, las cuales no favorecen en absoluto el aprendizaje.

Finalmente, el requisito *Req-03* trata de utilizar la propia actividad evaluadora para fomentar el aprendizaje. Para ello, no podremos asignar las tareas de evaluación de forma aleatoria, sino que debemos seguir unas pautas que aseguren que se fomenta el aprendizaje.

Este artículo presenta una metodología para diseñar y realizar actividades de evaluación por pares de forma que se satisfagan estos tres requisitos. Dicha técnica se ha aplicado durante dos años consecutivos a una asignatura de *Ingeniería del Software* en la cual se seguía una metodología de aprendizaje basada en proyectos [6]. Los datos recopilados tras su aplicación indican que dicha metodología alcanza los objetivos perseguidos.

Tras esta introducción, el resto de este trabajo se estructura como sigue: El apartado 2 describe el contexto en el cual se enmarca este trabajo. El apartado 3 explica la organización general de la metodología propuesta. El apartado 4 aborda el problema específico de asignar trabajos a alumnos para su evaluación. El apartado 5 muestra y analiza los resultados concretos empíricos

obtenidos. Finalmente, el apartado 6 sirve de sumario y cierre a este artículo.

2. Motivación: evaluación por pares en aprendizaje basado en proyectos

Esta metodología se utilizó dentro de una asignatura de *Ingeniería del Software* donde se seguía una *metodología de aprendizaje basada en proyectos* [6]. Si el objetivo de la asignatura era que los alumnos aprendiesen a desarrollar un proyecto software, entendíamos que la mejor manera de aprender era precisamente mediante el desarrollo de un proyecto software. El proyecto se desarrollaba por grupos. Cada grupo de trabajo se componía de 4 ó 5 alumnos.

Para el desarrollo del proyecto, se creó un portafolio electrónico [2] sobre la plataforma *Moodle*. La definición de las etapas de dicho portafolio fue fácil, pues cada etapa se correspondía con una de las fases del ciclo de vida software. Concretamente, la asignatura de *Ingeniería del Software I* cubría las siguientes etapas del ciclo de vida:

Idea de negocio Sobre una idea común propuesta por el equipo docente, cada grupo de alumnos tenía que elaborar su propia idea de negocio. Por ejemplo, durante el curso 2012-1023 los alumnos tuvieron que desarrollar plataformas *on-line* para la venta de bienes y servicios con grandes descuentos, al estilo de las plataformas comerciales *Groupo* o *Planeo*.

Especificación de requisitos. Una vez elaborada la idea de negocio, cada grupo debía producir una *especificación de requisitos* donde se detallaban tanto los requisitos funcionales como los no funcionales, los primeros de ellos modelados mediante *casos de uso*.

Diseño arquitectónico. A partir de la especificación de requisitos, los alumnos tenían que proporcionar el diseño de la arquitectura software del sistema, normalmente siguiendo un modelo arquitectónico de tres capas.

Diseño detallado. Por último, cada grupo debía proporcionar el diseño interno de varios de los componentes identificados a nivel arquitectónico, me-

dian *diagramas de clases, diagramas de actividades y máquinas de estado.*

Por cada etapa definida dentro de nuestro portafolio, los alumnos debían entregar un informe que recogiese los resultados de dicha etapa, así como algunas cuestiones destinadas a evaluar también el método de trabajo seguido dentro de dicha etapa. Para que la evaluación fuese efectiva, se debía realizar primero una *evaluación formativa* [3].

Dicha evaluación formativa debía informar al alumno de los errores cometidos, las razones por las que dichos errores se consideraban como tal, y una serie de indicaciones acerca de cómo corregirlos. Dado que: (1) los proyectos tenían un cierto tamaño; (2) las primeras versiones estaban plagadas de errores, como, por otro lado, era de esperar de alumnos que están aprendiendo; y (3) para que la evaluación fuese efectiva tenía que ser realizada en un corto periodo de tiempo; esta tarea de evaluación se nos antojaba faraónica.

Por ese motivo, decidimos implantar un sistema de *evaluación por pares*, el cual se describe a continuación.

3. Organización general del sistema de evaluación por pares

Con objeto de hacer la tarea de evaluación comentada en la sección anterior abordable, decidimos implantar un sistema de evaluación por pares. Dicho sistema se organizaba de la siguiente forma.

En primer lugar, dentro de cada etapa del portafolio, se realizaba un seguimiento o evaluación formativa informal del trabajo de los alumnos durante las sesiones de prácticas destinadas al desarrollo del proyecto. Dicha evaluación formativa informal, por razones obvias de tiempo, no era ni sistemática ni exhaustiva. Simplemente se centraba en orientar el trabajo de cada grupo respecto a aquellos aspectos que se consideraban más críticos. Por tanto, para asegurar el aprendizaje completo del alumnado, era necesario proporcionar mecanismos de evaluación más sistemáticos y exhaustivos. Para ello se utilizaba un mecanismo de entrega de informes en tres fases que describimos a continuación.

En primer lugar, para cada etapa del portafolio, cada grupo de alumnos entregaba una primera versión (provisional) del informe de resultados correspondiente a dicha etapa.

En una segunda etapa, cada informe entregado se asignaba para su evaluación a un grupo diferente al que elaboró dicho trabajo. Para asegurar que cada grupo era capaz de realizar la actividad evaluadora, por cada etapa del portafolio, se proporcionaba al alumnado una *rúbrica* [9] y un conjunto de instrucciones detalladas acerca de cómo realizar un proceso de evaluación

objetivo. Además, de forma complementaria, cada grupo podía proporcionar los comentarios y sugerencias, siempre constructivas, que considerase oportunos.

Una vez realizadas las evaluaciones, los informes de evaluación se enviaban a los grupos correspondientes, de forma que éstos pudiesen corregir los errores detectados. Tras corregir dichos errores, los informes definitivos se entregaban al equipo docente, que procedía a su evaluación, tanto formativa como sumativa.

Esta técnica, tal como se ha descrito hasta ahora, sólo permite asegurar que los alumnos están capacitados para realizar la tarea de evaluación. No obstante, para satisfacer el requisito *Req-01* completamente, debemos asegurar que el alumnado realiza la actividad evaluadora con la seriedad y el rigor que ésta merece.

Desgraciadamente la experiencia personal de los autores de este trabajo indica que para que el alumnado se tome una actividad en serio, dicha actividad debe tener algún tipo de peso, por mínimo que sea, en la calificación final de la asignatura. Por tanto, decidimos que la actividad evaluadora también necesitaba ser calificada.

Dado que cada grupo tenía que elaborar un *informe de evaluación*, decidimos calificar dicho informe. De esta forma, la calificación final de cada etapa del portafolio venía dictaminada por la media ponderada de las calificaciones del *informe de resultados* (90 %) y del *informe de evaluación* (10 %). De esta forma, asegurábamos que el alumnado realizaba la actividad con suficiente rigor y seriedad.

La calificación del *informe de evaluación* se realizaba de forma un poco subjetiva, ya que se trataba simplemente de verificar que la actividad se había realizado con seriedad y coherencia. Es decir, se trataba de corroborar, mediante técnicas de *perspective-based reading* [8], que los informes de evaluación entregados carecían de errores de bulto.

Respecto al requisito *Req-02*, entendemos que la tarea de evaluación es asumible para el alumno, ya que cada grupo (formado por 4 ó 5 alumnos) debía evaluar un único informe.

Por tanto, tan sólo restaba satisfacer el requisito *Req-03*, lo cual se conseguía mediante una estrategia de asignación de tareas de evaluación que se describe en el siguiente apartado.

4. Asignación de tareas para el fomento del aprendizaje

Para satisfacer el requisito *Req-03* ideamos una estrategia de asignación de informes de resultados a grupos para su posterior evaluación. Antes de mostrar dicha estrategia, abordaremos una serie de cuestiones preliminares que tuvimos que tener en consideración.

4.1. Mecanismos de aprendizaje en sistemas de evaluación por pares

Si queríamos potenciar el aprendizaje mediante sistemas de evaluación por pares, la primera tarea a realizar era analizar cómo se produce el aprendizaje mediante la utilización de estos sistemas. Concretamente, identificamos tres mecanismos diferentes de aprendizaje:

Identificación de errores propios Cada grupo, al recibir los resultados de la actividad evaluadora, adquiere conocimiento de sus errores y cómo solucionarlos, por lo que se produce un cierto aprendizaje.

Observación de errores de terceros Cada grupo, al evaluar el trabajo de otros, refuerza lo aprendido mediante la aplicación, desde la perspectiva del evaluador, de los conocimientos adquiridos. Además, mediante la evaluación del trabajo de terceros, aprende a identificar errores comunes y a cómo evitarlos.

Observación de aciertos de terceros La observación del trabajo de terceros permite descubrir nuevos puntos de vista y soluciones alternativas para un problema dado. De esta forma, el alumno puede tomar prestadas ingeniosas soluciones diseñadas por sus compañeros y adaptarlas a su propio trabajo.

Además, para evitar la sobrecarga de trabajo, debíamos analizar el esfuerzo que requiere realizar una actividad evaluadora. En este aspecto, cabe mencionar que es más fácil evaluar un trabajo excelente que un trabajo de escasa calidad. Sirva como ejemplo el trabajo de un revisor de un comité científico. En caso de tener que evaluar un trabajo de escasa calidad, aparte de rechazar el trabajo, dicho revisor, si realiza su tarea con diligencia, debe: (1) indicar los errores detectados; (2) justificar técnicamente dichos errores; y, en la medida de lo posible, (3) indicar posibles soluciones o alternativas. Sin embargo, si el trabajo a revisar fuese excelente, bastaría con felicitar a los autores y aceptarlo. Huelga decir que los trabajos excelentes se leen y asimilan más rápido que los trabajos malos, cuyos contenidos, por erróneos, son normalmente más difíciles de comprender.

Bajo estas suposiciones, nos percatamos de que si la asignación de los informes para la realización de la actividad evaluadora se hacía de forma aleatoria, evitando únicamente que un grupo evaluase su propio informe, podían darse los siguientes problemas:

1. Si un grupo con buenos resultados evalúa el trabajo de un grupo con muy malos resultados, en lugar de premiar al primer grupo por sus buenos resultados, le estábamos asignando una mayor carga

de trabajo. Por tanto, en este caso, estaríamos sobrecargando a un grupo que ya ha trabajado bien y bastante, lo que iría en contra del requisito *Req-02* (Ver Cuadro 1).

2. Además, si un grupo con buenos resultados evalúa el trabajo de un grupo con muy malos resultados, el primer grupo tendría pocas oportunidades de aprender por observación, lo cual iría en contra del requisito *Req-03* (Ver Cuadro 1).
3. Por otro lado, si un grupo con muy malos resultados revisaba el trabajo de un grupo con muy buenos resultados, obtendría una menor carga de trabajo como recompensa a sus pobres resultados, lo cual no nos parecía justo. Tampoco estábamos seguros de que el grupo evaluador estuviese plenamente capacitado para ejecutar la actividad evaluadora, lo cual iría en contra del requisito *Req-01* (Ver Cuadro 1).
4. Además, si un grupo con muy malos resultados revisaba el trabajo de un grupo con muy buenos resultados, el aprendizaje por observación podía degradarse hasta convertirse en un mero y burdo plagio de las soluciones propuestas por el grupo evaluado. El plagio, además de no conducir al aprendizaje, es moralmente y legalmente reprochable, iría en contra del requisito *Req-03* (Ver Cuadro 1)

Por tanto, decidimos que debíamos evitar asignaciones en las cuales hubiese una gran disparidad entre la calidad técnica del autor y del evaluador. Además, las asignaciones debían favorecer el aprendizaje por observación y facilitar la detección de errores. Para ello diseñamos la estrategia de asignación que se describe en el siguiente apartado.

4.2. Asignación de informes a grupos para su evaluación

El método diseñado funciona tal como se describe a continuación:

1. En primer lugar, definíamos unos niveles informales para clasificar cada trabajo. Por ejemplo, *bueno, medio y regular*.
2. A continuación, asignamos cada trabajo a un nivel diferente mediante una lectura superficial de cada trabajo y a partir de nuestras impresiones subjetivas sobre el rendimiento de cada grupo.
3. A continuación, por cada trabajo evaluábamos de forma rápida los aspectos positivos y negativos más prominentes de cada informe. Por ejemplo, para la etapa de especificación de requisitos, podríamos identificar aspectos como corrección de los diagramas de casos de uso o el análisis de requisitos no funcionales.

4. Finalmente, tratábamos de asignar trabajos a grupos para su evaluación, de forma que se satisficieran de la mejor forma posible una serie de restricciones que, a nuestro entender, ayudaban a fomentar el aprendizaje:

Restricción 1 Un trabajo sólo podía ser revisado por un grupo de un nivel adyacente o igual al nivel del grupo autor del informe. Por ejemplo, un informe de un grupo *bueno* sólo podría ser revisado por otro grupo *bueno* o por un grupo *medio*, pero nunca por un grupo *regular*. Esto contribuía a satisfacer los requisitos *Req-02* y *Req-03*.

Restricción 2 Al asignar un trabajo a un grupo, debíamos fomentar que los aspectos identificados como negativos en el informe a evaluar fuesen aspectos identificados como positivos en el informe entregado por el grupo evaluador. De esta forma aseguramos que el grupo evaluador sabe realizar dichos aspectos concretos, lo cual es un prerrequisito para poder evaluarlo. Esto contribuía a satisfacer el requisito *Req-01*.

Restricción 3 Al asignar un trabajo a un grupo, debíamos fomentar que algún aspecto identificado como negativo en el informe del grupo evaluador fuese un aspecto positivo del informe a evaluar. De esta forma potenciábamos que el grupo evaluador pudiese aprender por observación. Esto contribuía a satisfacer los requisitos *Req-01* y *Req-03*.

Restricción 4 Al asignar los trabajos, debíamos procurar que cada grupo pudiese corregir sus aspectos negativos bien mediante observación o porque se le notificasen los errores detectados. Esto contribuía a satisfacer el requisito *Req-03*.

Restricción 5 Para maximizar el aprendizaje, debíamos evitar que si el grupo G_i evaluaba al grupo G_j , entonces el grupo G_j evaluase al grupo G_i . Esto contribuía a satisfacer el requisito *Req-03*.

La última restricción no es trivial de entender, por lo que procedemos a explicarla detenidamente. Supongamos que tenemos tres grupos, a los que denominaremos G_i , G_j , G_k . Supongamos además que el grupo G_j presenta deficiencias en un aspecto A en el cual destacan tanto G_i como G_k .

Si G_j evalúa el trabajo del grupo G_k , el grupo G_j puede aprender por observación cómo mejorar el aspecto A . Si, además, el grupo G_k evalúa el trabajo del grupo G_j , el grupo G_j recibiría una evaluación formativa sobre el aspecto A cuyas conclusiones serían muy similares a las que ya habría alcanzado previamente el grupo G_j por observación.

	G1	G2	G3	G4	G5	G6
Identificación Requisitos	+	+	+	-	-	-
Casos de Uso	+	+	-	+	-	+
Plantillas Casos de Uso	-	-	-	-	+	-
Diagramas Secuencia	-	-	-	-	-	-
Req. No Funcionales	+	+	-	-	-	-
Nivel	B	B	M	M	M	R

B: Bueno; M: Medio; R: Regular

Cuadro 2: Evaluación informal previa

Por dicha razón tratamos de evitar el ciclo entre G_j y G_k . Para ello, haremos que el trabajo del grupo G_j lo revise el grupo G_i , que también destaca por el aspecto A . De esta forma, el grupo G_j puede recibir una evaluación formativa que incluya nuevos puntos de vista o aspectos complementarios a los adquiridos mediante la observación del trabajo del grupo G_k .

Dado que entender esta estrategia no es trivial y descrita en abstracto puede ser confusa, la siguiente sección la ilustra mediante un ejemplo concreto extraído de nuestra experiencia docente.

4.3. Ejemplo práctico: asignación de tareas de evaluación para un informe de requisitos

El ejemplo concreto que nos servirá para explicar la estrategia de asignación está extraído de una experiencia real perteneciente al curso 2011-2012. Dicho ejemplo se refiere a la asignación de tareas de evaluación de un *Informe de Especificación de Requisitos*, correspondiente a la segunda etapa de un proyecto de desarrollo software, de acuerdo al portafolio que nosotros definimos. El curso estaba compuesto por un total de 6 grupos de trabajo, que identificaremos, por motivos de privacidad, como $G1$ a $G6$.

Para dicho informe, se definieron como aspectos prominentes los siguientes elementos:

1. Identificación de requisitos, los cuales se recogían mediante lenguaje natural.
2. Modelado de los requisitos identificados mediante diagramas de casos de uso.
3. Especificación detallada de los casos de uso mediante plantillas¹.
4. Especificación detallada de ciertos casos de uso mediante diagramas de secuencia.
5. Identificación de requisitos no funcionales de relevancia.

El Cuadro 2 muestra los resultados de la evaluación superficial que es necesario realizar antes de la asigna-

¹Dependiendo de las características de cada caso de uso, cada grupo debía decidir si detallarlo por medio de plantillas o por medio de diagramas de secuencia.

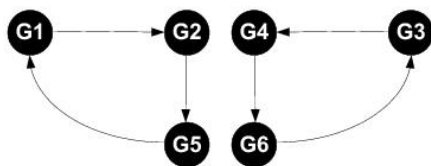


Figura 1: Solución para el ejemplo del Cuadro 2

ción de los trabajos. En las columnas, se muestran los diferentes grupos de trabajo. En las filas, a excepción de la última, cada uno de los aspectos prominentes identificados. La última fila sirve para asignar una categoría (*bueno (B)*, *medio (M)*, *regular (R)*) a cada trabajo. Cada celda (i, j) , a excepción de las pertenecientes a la última fila, contiene un + si el aspecto prominente de la fila i ha sido satisfactoriamente realizado por el grupo de la columna j . Si ha sido insatisfactoriamente realizado, dicha celda contendrá un -. Si dicho aspecto simplemente presenta un nivel cuasi aceptable, la celda aparece vacía.

La Figura 1 muestra una posible asignación, en forma de grafo dirigido, que satisface las restricciones descritas en la sección anterior. En este grafo, la existencia de un arco entre los nodos G_i y G_j indica que el grupo G_i debe revisar el trabajo del grupo G_j .

Explicamos brevemente parte de dicha solución. Partimos del grupo $G1$, que es el que mejores resultados presenta. Dicho grupo debería tener una tarea de evaluación liviana, en premio al esfuerzo previo realizado. Por ello podría revisar al grupo $G2$ y corregir algunos de los errores detectados en este grupo relativos a la realización de las plantillas de casos de uso.

El grupo $G2$ posee un rendimiento bueno, pero inferior al del grupo $G1$. En concreto, presentaba ciertas deficiencias con respecto a la realización de las plantillas de casos de uso. Por otra parte, el grupo $G5$ destacaba en dicho apartado. Por tanto, si el grupo $G2$ evaluaba el trabajo de $G5$ podría aprender por observación. El grupo $G2$, además, destacaba por la tarea de identificación de requisitos, aspecto débil del grupo $G5$. Por tanto, el grupo $G2$, al revisar el trabajo del grupo $G5$ podría aportarle valiosas sugerencias. Por tanto, asignamos el grupo $G5$ al grupo $G2$.

El grupo $G5$ destacaba por la realización de las plantillas de casos de uso, aspecto mejorable del grupo $G1$. Por tanto, si $G5$ evaluaba a $G1$, $G5$ podría contribuir con comentarios y sugerencias a mejorar el trabajo de $G1$. Además, $G5$ podía, por observación, aprender a mejorar ciertos aspectos de su propio trabajo, como la identificación de requisitos, aspecto destacado de $G1$. Por tanto, decidimos que el grupo $G5$ evaluase el trabajo del grupo $G1$.

Adviértase en este caso el problema de los ciclos abordado por la *Restricción 5*. Al asignar los trabajos,

podíamos haber decidido que el grupo $G2$ evaluase al grupo $G5$ y el grupo $G5$ al grupo $G2$, ya que había buenas razones para realizar ambas asignaciones.

No obstante, si evitamos este ciclo, el grupo $G5$ recibiría un doble refuerzo en el apartado de *Identificación de Requisitos*. Por una parte, el grupo $G5$ podía aprender por observación mediante la evaluación del grupo $G1$; pero también podía aprender a partir de la evaluación formativa recibida del grupo $G2$. Este doble refuerzo no se daría si hubiésemos creado un ciclo entre $G2$ y $G5$.

Un procedimiento similar se aplicaría a los grupos $G3$, $G4$ y $G6$.

La siguiente sección describe los resultados empíricos recogidos tras aplicar nuestra metodología a diversos cursos de *Ingeniería del Software I*.

5. Análisis de los resultados obtenidos

Esta sección muestra y analiza los datos empíricos recopilados a través de una serie de encuestas realizadas al alumnado al finalizar las asignaturas donde habíamos aplicado la metodología presentada en este artículo. La metodología se aplicó a tres cursos de *Ingeniería del Software I*. Dos de ellos, de 9 créditos, pertenecían a la titulación a extinguir de *Ingeniería Informática*, se impartieron en los cursos académicos 2011-2012 y 2012-2013, y tuvieron 28 y 31 alumnos matriculados, respectivamente. El tercer curso, de 6 créditos ECTS (*European Credit Transfer System*), pertenecía a la titulación de nueva implantación *Grado en Ingeniería Informática*, se impartió durante el curso 2012-2013; y tenía 35 alumnos matriculados.

Las encuestas realizadas estaban destinadas a verificar si la metodología satisfacía los objetivos planteados. Dado que entendíamos que preguntar a los alumnos si el sistema les sobrecargaba era ocioso, ya que la respuesta sería "sí"; lo que debíamos comprobar era, básicamente, si realmente el sistema contribuía al aprendizaje. Para ello debíamos comprobar tanto si las evaluaciones recibidas les eran de utilidad como si el hecho de evaluar a otros les ayudaba a aprender.

Respecto a la sobrecarga del alumnado, entendíamos que tener que evaluar un informe en el plazo de una semana, pudiendo dividir el trabajo entre 4 personas, no era una carga de trabajo excesiva.

Las encuestas realizadas tenían un total de 60 preguntas que pretendían analizar diferentes aspectos de la asignatura, así como tratar de verificar la consistencia de las respuestas, con el objetivo de evitar que se contestase al azar. En este artículo sólo mostramos las respuestas que conciernen al objetivo del mismo, es decir, al sistema de evaluación por pares.

	2011-2012		2012-2013 (Ing)		2012-2013 (Grado)	
	23 encuestas		13 encuestas		9 encuestas	
Pregunta	Media	Desv.	Media	Desv.	Media	Desv.
Revisar el trabajo de terceros me ha servido para aprender	3.46	0.60	3.00	1.11	3.44	1.03
Revisar el trabajo de terceros me ha servido para mejorar	3.65	0.58	3.23	1.09	3.78	0.83
Las correcciones recibidas me han servido para mejorar	3.42	0.72	3.38	0.96	3.44	0.88

Cuadro 3: Resultados del test del rendimiento de la estrategia aplicada

El Cuadro 3 muestra dichas preguntas y los resultados obtenidos para cada curso. Cada pregunta se contestaba con valores de una escala entre 1 y 5 donde 1 significaba *nada* y 5 *mucho*. Junto a cada pregunta se muestra el valor medio y desviación de las respuestas recopiladas.

Las encuestas las completaron un total de 23 alumnos en el curso 2011-2012, 13 alumnos de *Ingeniería Informática* en el curso 2012-2013 y 9 alumnos de *Grado en Ingeniería Informática* en el curso 2012-2013. La razón del decremento en el número de encuestas realizadas en el curso 2012-2013 radica en el hecho de que mientras en el curso 2011-2012 dábamos 0.25 puntos extra en el examen al alumno que completase la encuesta, durante el curso 2012-2013 eliminamos dicha recompensa. Esto parece confirmar nuestra hipótesis de que de que la mayoría de los alumnos sólo realizan aquellas actividades que tienen repercusión en su calificación final.

Como puede observarse, los resultados son prometedores, y las medias suelen estar por encima del valor considerado como *Normal*, siendo tendentes al valor *Bastante*. Por tanto, entendemos que un amplio sector del alumnado considera que la metodología contribuye al aprendizaje. Los valores altos en la desviación para el curso 2012-2013 la atribuimos al hecho de corresponder a una población menor de datos. En este caso, las respuestas de alumnos muy insatisfechos, que evalúan sistemáticamente todas las preguntas de la encuesta con 1 ó 2, contribuyen a elevar el valor de la desviación.

En un aspecto menos cuantitativo, nuestra experiencia personal es que la metodología nos descarga de trabajo y contribuye, de algún u otro modo, a mejorar el rendimiento de los grupos de alumnos. Las versiones definitivas de los informes suelen ser mejores que las iniciales, y ello se debe al sistema de evaluación por pares. Además, al recibir nosotros informes con menos errores, nos vemos descargados de trabajo, pues son informes que se pueden evaluar de manera formativa de forma más rápida, pues no hay que notificar ni justificar tantos errores.

No se ofrecen datos cuantitativos relativos a la mejora de la tasa de éxito de los cursos donde se ha aplicado esta metodología en comparación con cursos anterior-

res donde no se ha aplicado. La razón es que no podemos considerar los datos relativos a cursos anteriores ala metodología como fiables. En dichos cursos los autores de este trabajo no impartían docencia, por lo que el sistema era diferente. La tasa de éxito en curso anteriores estaba en torno al 100% de los matriculados pero una simple interacción con los alumnos permitía comprobar que la mayoría de ellos no alcanzaban los objetivos de aprendizaje, por lo que consideramos dicho dato como no fiable.

Un aspecto controvertido de nuestro trabajo y que suele levantar cierta sorpresa, y por tanto merece la pena analizar, es el hecho de que los grupos con peor rendimiento se evalúen entre ellos. El lector podría argumentar que este sistema no favorece el aprendizaje de dichos grupos. Nuestra experiencia revela que sí.

La razón principal es la siguiente: Si un grupo con mal rendimiento recibe un informe de evaluación de un grupo con buen rendimiento, recibirá una larga lista de aspectos a mejorar, muchos de los cuales ni siquiera comprenderá. Ello normalmente desmotiva o desalienta al grupo con mal rendimiento, que puede incluso llegar a abandonar la asignatura. Sin embargo, un informe de un grupo con un rendimiento medio o malo, contendrá menos sugerencias de mejora, pero el esfuerzo global para el grupo que recibe el informe será más asequible, por lo cual será más propenso a realizar dichas mejoras. Es decir, se trata de incentivar a mejorar más que desanimar. Por tanto, entendemos que la revisión por parte de grupos con un rendimiento similar ayuda a mejorar a base de pequeños incrementos.

Finalmente, nos gustaría analizar un último aspecto que consideramos relevante. El lector podría argumentar, no sin cierta lógica, que la metodología que exponemos es compleja y, por tanto, el esfuerzo que hay que realizar para llevarla a cabo podría ser igual, o superior incluso, al que tendríamos que realizar si evaluásemos cada grupo de alumnos nosotros mismos directamente.

Aunque la metodología es compleja, su tiempo de ejecución para un número de grupos no muy alto, entre 6 y 10 grupos, no es elevado; y es notablemente inferior al necesario para evaluar con carácter formativo todos los informes entregados. Por tanto, la metodología sí nos está ayudando a salvar cierto tiempo de trabajo, a la par que parece contribuir al aprendizaje.

Señalar, por último, que la parte de la metodología que más tiempo requiere es la aplicación de la estrategia de asignación. En este sentido, estamos trabajando para crear un *plug-in* para la plataforma *Moodle* que encuentre una asignación óptima automáticamente. Para ello, utilizando como entrada una tabla como la del Cuadro 2, se plantearía un problema de satisfacción de restricciones, el cual se resolvería automáticamente por medio de un computador.

6. Sumario y trabajos futuros

Este artículo ha presentado una metodología para organizar sistemas de evaluación entre pares de forma que: (1) se disminuya la carga de trabajo de los equipos docentes, permitiendo sistemas de evaluación continua más razonables y escalables; (2) no se sobrecargue en exceso al alumnado; y (3) contribuya al aprendizaje.

Diseñar dicha metodología no fue en absoluto trivial, pero una vez ideada resultó relativamente sencilla y rápida de aplicar. Además, según hemos podido comprobar, ayuda a alcanzar los objetivos que perseguíamos: (1) nuestra carga de trabajo se vio aliviada; y, (2) los trabajos de cada grupo, salvo inevitables excepciones, mejoraron.

La metodología permite asegurar que la actividad evaluadora se realiza con cierta seriedad y rigor a la par que fomenta el aprendizaje. Este último punto se satisface mediante un elaborado sistema de asignación de informes a grupos para su evaluación que permite asegurar que: (1) la carga de trabajo que reciben los grupos es proporcional al esfuerzo realizado; (2) cada grupo recibe una adecuada evaluación formativa; (3) se fomenta el aprendizaje por observación; y (4) se evitan plagios indeseados.

Dicha metodología se ha aplicado con cierto éxito a tres cursos de *Ingeniería del Software*, pertenecientes a las titulaciones de *Ingeniería Informática* y *Grado en Ingeniería Informática*.

Como trabajo futuro, nuestra intención es proporcionar un soporte informático que permita la automatización de parte de la metodología, en particular de la parte relativa a la asignación de informes a evaluar. Nuestro objetivo último es integrar dicho soporte en diversas plataformas de aprendizaje, y en particular en *Moodle*, que es nuestra plataforma de trabajo habitual.

Agradecimientos

Este trabajo se encuadra dentro del proyecto de innovación docente “*Aprendizaje basado en proyectos para Ingeniería del Software*”, financiado por la Univer-

sidad de Cantabria. Nos gustaría agradecer también a Daniel Sadornil, sus valiosos y bien intencionados comentarios.

Referencias

- [1] Charles C. Bonwell and James A. Eison. *Active Learning: Creating Excitement in the Classroom*. Jossey-Bass, Febrero 1991.
- [2] Ann Bullock and Parmalee P. Hawk. *Developing a Teaching Portfolio: A Guide to Preservice and Practicing Teachers*. Prentice Hall, Julio 2004.
- [3] Kay Burke. *Balanced Assessment: From Formative to Summative*. Solution Tree, Febrero 2010.
- [4] María Cavas, Francisco Chicano, Francisco Luna, and Luis Molina. Autoevaluación y Evaluación entre Iguales en una Asignatura de Redes de Ordenadores. In *Actas de las XVII Jornadas de Enseñanza Universitaria de la Informática (JENUI)*, pages 303–310, Sevilla (España), Julio 2011.
- [5] Tim S. Roberts. *Self, Peer and Group Assessment in E-learning*. Information Science Publishing, Enero 2006.
- [6] Pablo Sánchez and Carlos Blanco. Experiencias y Lecciones Aprendidas de la Implantación de una Metodología de Aprendizaje basada en Proyectos para una Asignatura de Ingeniería del Software I. In *Actas de las XVIII Jornadas de Enseñanza Universitaria de la Informática*, pages 41–48, Ciudad Real (España), Julio 2012.
- [7] Jesús Serrano-Guerrero, Francisco P. Romero, Emilio Fdez-Viñas, and José A. Olivás. La sobre-evaluación. In *Actas de las XVII Jornadas de Enseñanza Universitaria de la Informática (JENUI)*, pages 91–100, Sevilla (España), Julio 2011.
- [8] Forrest Shull, Ioana Rus, and Victor R. Basili. How perspective-based reading can improve requirements inspections. *IEEE Computer*, 33(7):73–79, Julio 2000.
- [9] Dannelle D. Stevens and Antonia J. Levi. *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback and Promote Student Learning*. Stylus Publishing, Noviembre 2004.
- [10] Miguel Valero-García and Luís M. Díaz de Cerio. Evaluación Continuada a un Coste Razonable. In *Actas de las IX Jornadas de Enseñanza Universitaria de la Informática (JENUI)*, pages 183–190, Cádiz, (España), Julio 2003.