



UNIVERSIDAD
COMPLUTENSE
MADRID

Proyecto de Innovación y Mejora de la Calidad Docente

PIMCD 91/2014

Implementación y puesta en práctica de técnicas de
evaluación masiva, evaluación por pares y
autoevaluación para MOOCs.

Carlos Gregorio Rodríguez

Facultad de Ciencias Matemáticas

Departamento de Sistemas Informáticos y Computación

16 de julio de 2015

Resumen

Entre los nuevos retos que tiene que afrontar la Universidad están los cambios de paradigmas educativos que propician las tecnologías de la información.

Los dispositivos y redes que actualmente permiten la comunicación a Internet de forma continua y ubicua a miles de millones de personas, posibilitan la existencia de los MOOC (Massive Open Online Courses). La Universidad tiene sin duda que considerar este escenario, en el que millones de personas pueden llamar a su puerta a interesarse por todo aquello que tiene que ofrecer, y tomar las medidas adecuadas para prepararse para un futuro inminente.

Como todo proceso incipiente y en desarrollo, la planificación y creación de MOOCs plantea numerosas incógnitas que habrá que ir estudiando y para las que habrá que buscar soluciones.

Uno de los grandes retos que plantea el paradigma educativo de los MOOCs es la evaluación. Este proyecto propone hacer una reflexión y una puesta en práctica experimental de diferentes modelos de evaluación (autoevaluación, evaluación continua y evaluación por pares) adecuados para cursos con elevado número de participantes.

Índice

| | |
|--|-----------|
| 1. Objetivos propuestos en la presentación del proyecto | 2 |
| 2. Objetivos alcanzados | 3 |
| 3. Metodología empleada en el proyecto | 6 |
| 3.1. Metodología de trabajo | 6 |
| 3.2. Metodología de desarrollo de software | 6 |
| 3.3. Metodología del experimento | 6 |
| 4. Recursos humanos | 7 |
| 5. Desarrollo de las actividades | 8 |
| 6. Conclusiones | 12 |

1. Objetivos propuestos en la presentación del proyecto

Si bien las metodologías educativas están cambiando –como demuestran, por ejemplo, los proyectos presentados en las convocatorias de PIMCD de la UCM– y es fácil encontrar experiencias de innovación en materiales, en didácticas, en herramientas, etc., los métodos de evaluación, por el contrario, evolucionan mucho más lentamente.

Sin embargo, son los métodos de evaluación uno de los aspectos pedagógicos en los que más se fijan los estudiantes. No sin razón, pues son ellos los que finalmente tienen que someterse a dicha evaluación y los que van a tener que enfrentar el éxito o el fracaso en dichas pruebas.

El profesorado que promueve este proyecto sigue varias directrices esenciales a la hora de diseñar en los objetivos de la formación, que son precisamente los que guían la evaluación:

- El acceso a la información no es un problema. No es necesario conocer datos e información de memoria. Contamos con conexión casi constante y ubicua a Internet y a otras personas. Un objetivo esencial de la educación es, por tanto, aprender a utilizar y a discriminar la información.
- En un mundo que cambia tan rápidamente es esencial formar a las personas para que puedan continuar aprendiendo. Autoaprendizaje y autoevaluación son habilidades esenciales, que están intrínsecamente ligadas, y que debemos fomentar.
- El futuro es desconocido y los retos y problemas que tendrán que enfrentar nuestros estudiantes en ese futuro necesitarán también de nuevas soluciones. Consideramos imprescindible fomentar la capacidad de síntesis y creatividad para generar soluciones a problemas y la capacidad de comparación y evaluación entre distintas alternativas.
- Los aspectos sociales y grupales ocupan cada vez más espacio en nuestra realidad. La cooperación y colaboración entre personas es algo natural y es el modo en el que la ciencia y la cultura en sus sentidos más amplios se construyen. El aprendizaje colaborativo es aún un área poco explorada que consideramos tiene que fomentarse y estudiarse.

Con este proyecto nos proponemos pensar sobre los métodos de evaluación y aprovechar el surgimiento de los MOOC para plantear estrategias que se adapten a este nuevo contexto educativo. El problema que plantea la evaluación de miles de participantes en un curso no puede resolverse con un simple test estándar online.

Nuestro proyecto quiere promover y experimentar estos nuevos modelos de evaluación que están íntimamente ligados a las herramientas informáticas que darán soporte a dichos métodos innovadores. Por ejemplo, igual que cada vez la navegación en la red es más particular para cada usuario, se puede pensar en técnicas de evaluación que sean adaptables a los distintos perfiles y recorridos de los participantes en los cursos, en particular en los MOOC; se pueden implementar métodos que permitan autoevaluaciones personalizadas en función del progreso en el curso; se puede fomentar conceptos tan importantes como colaboración, capacidad crítica y responsabilidad con buenas metodologías de evaluación por pares; podemos utilizar la versatilidad y potencia de los actuales sistemas informáticos, y de los desarrollos en inteligencia artificial, para proponer técnicas y tecnologías que permitan una evaluación lo más abierta posible, de tal forma que no haya una única respuesta correcta...

Teniendo en cuenta estas consideraciones, los objetivos planteados en el proyecto de innovación educativa son:

- Realizar un estudio del estado del arte en técnicas y metodologías de evaluación en las principales plataformas de MOOCs (Coursera, edX, Udacity, MiriadaX, Udemy...).
- Reflexionar sobre la adecuación de las distintas alternativas de evaluación a diferentes estudios o materias así como los requisitos técnicos necesarios para llevarlas a cabo.

- Planificar y adecuar unas estrategias de evaluación que permitan experimentar con técnicas de evaluación masiva, evaluación por pares y autoevaluación para la asignatura de Informática de primer curso en los grados ofertados en la Facultad de Matemáticas.
- Implementar e integrar en módulos de Moodle las herramientas adecuadas para llevar a cabo estas evaluaciones.

2. Objetivos alcanzados

Gracias a la generosa implicación de la mayoría de los integrantes del proyecto, éste ha podido desarrollarse en bastante profundidad y hemos podido diseñar y experimentar diferentes técnicas de seguimiento y evaluación. Concretando, algunos de los objetivos más destacables desarrollados para el primer curso de Informática en los grados de matemáticas son:

- Hemos incorporado técnicas denominadas de *rastreo y seguimiento* al diseño de las actividades y materiales que se incorporan en el campus virtual de la asignatura. De esta forma, la asignatura se va *desplegando* de forma particular para cada alumno en función de su trabajo y de sus resultados. Para poner un ejemplo muy sencillo, se puede hacer que una determinada actividad, digamos una entrega de prácticas, no sea posible, hasta que el alumno haya completado otras partes del curso: realizado ejercicios, cuestionarios, consultado materiales, etc. En la figura 1 puede verse una de las muchas formas en las que Moodle puede utilizar esa información, por ejemplo para mostrar un listado de actividades completadas por alumno.

Grupos visibles: Grupo B

Nombre: Todos A B C D E F G H I J K L M N Ñ O P Q R S T U V W X Y Z
 Apellido(s): Todos A B C D E F G H I J K L M N Ñ O P Q R S T U V W X Y Z
 Página: 1 2 3 (Siguiente)

| Nombre / Apellido(s) | Dirección de correo | Evaluación del desarrollo... | Calificaciones del examen... | Calificaciones del examen... | Calificaciones de las... | Sesión interactiva T14 | Sesión interactiva T15 | Sesión interactiva T16 | Soluciones a la hoja 7 de... | Practica obligatoria 4 | Entrega de la practica... | Sesión interactiva T10 | Video-Tutoriales Funciones | descomposicion.py | descomposicion.py | Sesión interactiva T11 | Hoja 4 de ejercicios | Soluciones a los ejercicios... | Reparso de operaciones con ... | Entrega de la practica... | Sesión interactiva T13 | Hoja 5 de ejercicios | de ordenación de... | e algoritmos... | n Turtle |
|-----------------------------|------------------------------|-------------------------------------|-------------------------------------|------------------------------|--------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Elisabeth Belmonte Martínez | elisabeth_totana@hotmail.com | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| Alma Mª Bermudo Bazo | abbr_100@hotmail.com | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| Roberto Bolaños García | rober.b.garcia.1a@gmail.com | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| antonio borrego | antborre@ucm.es | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| Lucia Canitrot | luciacan@ucm.es | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |

Figura 1: Vista de las actividades completadas

- Hemos utilizado cuestionarios de autoevaluación interactivos en los que los participantes pueden ir comprobando su comprensión de los aspectos más básicos de la asignatura.
- Uno de los aspectos más destacados ha sido introducir en la asignatura la evaluación por pares de ejercicios prácticos. Esta técnica de evaluación es la que ha resultado más costosa y exigente para los integrantes del proyecto. Cada vez que se diseñaba una evaluación por pares había que:
 - definir una rúbrica precisa que considerara la mayor parte de la casuística posible;

- proporcionar ejemplos de evaluación para que los participantes fueran adquiriendo unos criterios comunes;
- evaluar las prácticas con los mismos criterios que los participantes en el curso para así poder tener una evaluación *objetiva* con la que comparar las evaluaciones de los alumnos, y poder proporcionar así una calificación a la evaluación de los pares.

En la figura 2 puede verse una de las muchas vistas de la herramienta. En ella aparece cada participante con las evaluaciones recibidas, las evaluaciones realizadas (o no realizadas), cada una de las notas recibidas, cada una de las calificaciones otorgadas, las calificaciones finales por entrega y por evaluación de trabajos, las notas obtenidas por las herramientas automáticas (segunda columna después del nombre), etc.

| | uims user | uims user | uims user | | | | |
|--------------------------------|------------|-----------|-----------|---|----|--|----|
| Juan Manuel Espinosa Rodríguez | Práctica 4 | -1 | 0.00 | 71 (-) @ 0- JORGE CARMONA RUBER 0 (0) @ 0- Jorge Fernández Calatrava 71 (20) @ 0- Celia Fernández Martin 71 (20) @ 0- Sergio Sánchez | 71 | 50 (19) @ 0- Gonzalo Herrero Garcia 67 (11) @ 0- David Llorente Silveiras 67 (11) @ 0- Paula Sanz Chamorro | 14 |
| Laura Etayo Rodríguez | Práctica 4 | 37 | 0.00 | 30 (-) @ 0- Carlos Gregorio Rodriguez 33 (19) @ 0- Almudena Jaén Camarero 16 (11) @ 0- Tsvetomira Marinova 43 (11) @ 0- María Mora Jiménez | 30 | 50 (20) @ 0- Carlos Arias 47 (18) @ 0- Diego Castelló Moreno 50 (8) @ 0- María Mora Jiménez | 15 |
| Javier Feito | Práctica 4 | 11 | 0.00 | 61 (-) @ 0- JORGE CARMONA RUBER 58 (19) @ 0- Brigytte Daza Vega 74 (11) @ 0- Esther Fernández Moreno 71 (14) @ 0- Gloria Vázquez Vidal | 61 | 80 (11) @ 0- Esther Adán Vilares 71 (14) @ 0- Natalia Gordo Herrera 44 (4) @ 0- José Miguel Herrera Peinado | 9 |
| Jorge Fernández Calatrava | Práctica_4 | 100 | 0.00 | 80 (-) @ 0- JORGE CARMONA RUBER 67 (11) @ 0- Itziar Garcia Calle 67 (11) @ 0- Pedro González Garcia 80 (20) @ 0- Susana Sánchez Callejas | 80 | 0 (0) @ 0- Juan Manuel Espinosa Rodriguez 36 (20) @ 0- Carlos Martin Lorenzo 44 (20) @ 0- Juan Carlos Torres Navarro | 13 |
| Celia Fernández Martin | Práctica 4 | 0 | 0.00 | 67 (-) @ 0- JORGE CARMONA RUBER 58 (15) @ 0- Guillermo Barajas Ayuso - (0) @ 0- Silvia De Pablo Ordiales 44 (4) @ 0- Alberto Rey Abelaيرا | 67 | 34 (6) @ 0- Elena Carvajal Luque 71 (20) @ 0- Juan Manuel Espinosa Rodriguez 53 (11) @ 0- Carmen Raposo Jiménez | 12 |
| Esther Fernández Moreno | Práctica 4 | 26 | 0.00 | 61 (-) @ 0- JORGE CARMONA RUBER 47 (11) @ 0- Ari Eidelman Bokler 77 (8) @ 0- Bernardo Perea Foseca 77 (8) @ 0- Sergio Sánchez | 61 | 64 (6) @ 0- Beatriz Arruga Pérez 74 (11) @ 0- Javier Feito 49 (4) @ 0- marta saez moreno | 7 |
| Victor Fernández Sánchez | Práctica_4 | 0 | 0.00 | 44 (-) @ 0- JORGE CARMONA RUBER - (0) @ 0- Katherine Michelle Guachamin Rendón 31 (11) @ 0- Ana Ramos Santamaria 18 (2) @ 0- Anthony Rivera Urbina | 44 | 27 (7) @ 0- Javier Cobos del Álamo 41 (4) @ 0- Tsvetomira Marinova 53 (14) @ 0- david prieto casero | 8 |

Figura 2: Vista de la actividad Workprog desarrollada e integrada en Moodle

- También hemos podido adaptar e integrar en el desarrollo de la asignatura una herramienta de corrección automática, desarrollada en anteriores proyectos de innovación (PIMCD 33/2007 y PIMCD 220/2013), denominada FLOP. Esta herramienta proporciona una evaluación automática para ejercicios de programación.
- Todos estos aspectos comentados anteriormente han sido adaptados e integrados en la plataforma Moodle en la que se ha desarrollado el campus virtual para la asignatura de Informática, común en el primer curso de todos los grados impartidos en la Facultad de Matemáticas. Para ello, hemos exprimido las posibilidades más avanzadas del propio gestor de cursos Moodle y, en otras ocasiones, hemos tenido que desarrollar las aplicaciones informáticas necesarias e integrarlas.
- Por último, un aspecto que no estaba inicialmente contemplado en el proyecto y que también hemos tenido que considerar, y en cierta forma se mantiene dentro del espíritu del proyecto y lo completa aún más, es el uso de una herramienta automática de detección de plagios. La herramienta considerada se denomina MOSS y ha sido desarrollada en la Universidad de Stanford (<http://theory.stanford.edu/~aiken/moss/>). Esta herramienta a ayudado a detectar exitosamente, y evaluar adecuadamente, los plagios en las entregas

obligatorias de prácticas de programación. En las figuras 3 y 4 pueden verse ejemplos de uso de dicha herramienta. En la primera de ellas vemos el porcentaje de similitud entre trabajos y el número de líneas que comparten. En la figura 4 vemos la comparación de dos trabajos concretos, con distintos códigos de color nos muestran las zonas que ha identificado como muy similares.

| File 1 | File 2 | Línes Matched |
|---|--|---------------|
| martasaezmoreno/practica_2.py (81%) | RocAoSalgadoSoneira/practica_2.py (80%) | 141 |
| AlexMiltchev/practica2.py (75%) | FernandoPeA+aVÁizquez/Historam.py (79%) | 66 |
| AlbaRoldÁinOrejÁn/practica2.py (65%) | pilardeCastroZea/practica2.py (70%) | 91 |
| IvanSanchezGuio/practica_2.py (65%) | JavierFeito/practica_2.py (66%) | 56 |
| JosÁ@ManuelSandoval/PrÁctica2.py (61%) | MarÁaPiresChoupina/practica2.py (64%) | 57 |
| AlmudenaJaÁ@nCamarero/practica_2.py (67%) | ChristianLÁpezVicente/practica_2.py (61%) | 53 |
| AlexMiltchev/practica2.py (52%) | GonzaloHerreroGarcÁa/Histograma.py (45%) | 46 |
| AndreadelaFuenteSilva/Practica2.py (37%) | MarÁaCristinaPÁ@rezTorres/practica2.py (49%) | 64 |
| SaraArribasMartin/Practica2.py (61%) | XIAOCHENG/practica2.py (65%) | 67 |
| CristinaGarcÁaAdeleTorreOlÁas/practica_2.py (48%) | InÁ@sGarcÁ-aMorales/practica_2.py (53%) | 52 |
| FernandoPeA+aVÁizquez/Historam.py (47%) | GonzaloHerreroGarcÁa/Histograma.py (39%) | 43 |
| AlmaMÁBermudoBazo/practica2.py (46%) | SilviaDePabloOrdiales/Practica2.py (51%) | 45 |
| joshualopezsanchez/practica2.py (80%) | robertogarciaidiaz/practica2.py (80%) | 43 |
| GuillermoDelgadoPÁ@rez/Practica2.py (37%) | JuanManuelEspinosaRodriguez/Practica2.py (40%) | 37 |
| joshualopezsanchez/practica2.py (73%) | MarcosGonzÁilezSÁichez/practica2.py (74%) | 39 |
| MarcosGonzÁilezSÁichez/practica2.py (74%) | robertogarciaidiaz/practica2.py (73%) | 39 |
| BernardoPereaFoseca/Practica2.py (34%) | victorocÁaasantiago/practica2.py (47%) | 38 |
| ItziarGarciaCalle/practica2.py (46%) | NataliaGordoHerrera/practica2.py (42%) | 32 |
| NataliaGordoHerrera/practica2.py (42%) | SilviaDePabloOrdiales/Practica2.py (42%) | 31 |
| AlmaMÁBermudoBazo/practica2.py (37%) | NataliaGordoHerrera/practica2.py (41%) | 36 |

Figura 3: Informe del detector de plagios Moss

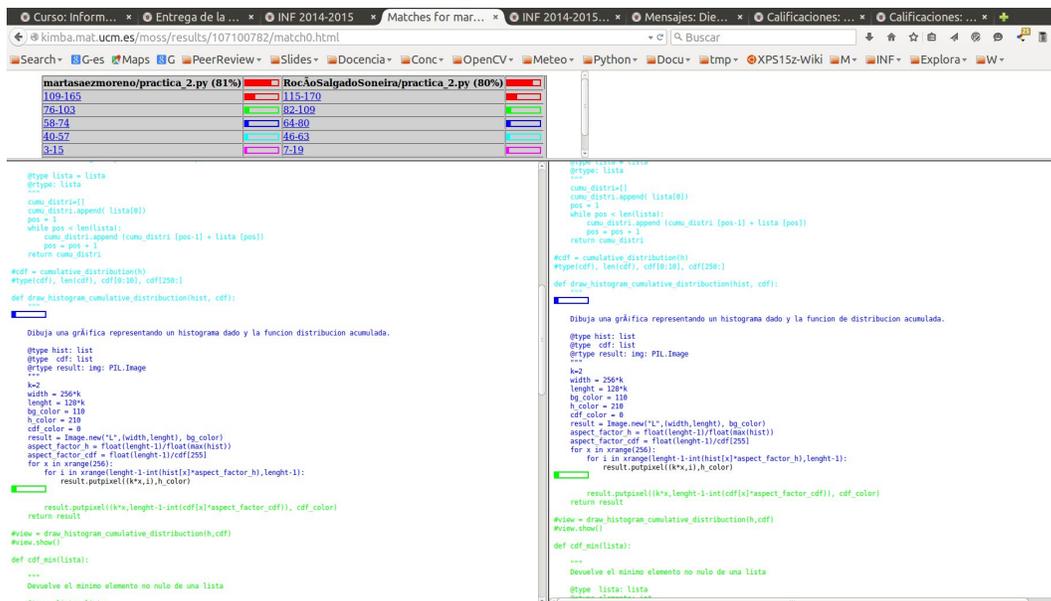


Figura 4: Detalle de la comparación de dos trabajos

3. Metodología empleada en el proyecto

Metodología es una palabra amplia y que se puede aplicar a múltiples facetas. Comentamos brevemente las metodologías más importantes relacionadas con el desarrollo del proyecto.

3.1. Metodología de trabajo

El equipo de profesores que integra el proyecto tiene una amplia experiencia en trabajo conjunto, tanto en investigación como en innovación educativa. La metodología de trabajo que habitualmente utilizamos es muy horizontal y colaborativa. Aparte de reuniones en las que consideramos y discutimos los diferentes aspectos para decidir las líneas a seguir, utilizamos repositorios compartidos con control de versiones en los que el trabajo de todos se va integrando y en el que cualquier participante pueden corregir, mejorar o completar el trabajo de los demás.

3.2. Metodología de desarrollo de software

Con respecto a la metodología de desarrollo de las herramientas software, también contamos con una nutrida experiencia y todos los integrantes hemos desarrollado con anterioridad herramientas para su uso tanto en docencia como en investigación, algunas de ellas con un uso muy continuado.

Una metodología que nos parece muy adecuada para el desarrollo de herramientas es la de prototipado, que consiste en desarrollar modelos de las aplicaciones de software básicos que se van enriqueciendo a medida que se va experimentando con ellos. La metodología de prototipado es además muy adecuada para el trabajo en grupos pequeños.

3.3. Metodología del experimento

Para poder evaluar de forma rigurosa la experiencia con las diferentes formas de evaluación, hemos diseñado, desde el punto de vista estadístico un experimento.

La Facultad de Matemáticas ofrece 5 grupos de primero, 4 de mañana y uno de tarde. Todos los grupos de primero están estrechamente coordinados en la asignatura de Informática desde hace años. Los contenidos, pruebas de evaluación intermedia (entrega de prácticas de programación) y exámenes son comunes a todos los grupos.

Para realizar el experimento hemos elegido los 4 grupos de la mañana por su homogeneidad en cuanto a tamaño y perfil del alumnado. De estos 4 grupos uno de ellos se ha dejado como grupo de control. En el grupo de control, el profesor ha realizado básicamente el mismo trabajo que en cursos anteriores. En los otros tres grupos hemos introducido en el campus virtual todas las actividades comentadas anteriormente, en particular aquellas detalladas en el apartado 2.

Todos los alumnos de los 4 grupos de la mañana realizaron un cuestionario inicial para identificar conocimientos y habilidades iniciales. También un cuestionario intermedio y un cuestionario final sobre el desarrollo de la asignatura. Los 3 grupos del experimento tenían alguna pregunta más que el grupo de control sobre aspectos concretos de las técnicas utilizadas.

Todos estos cuestionarios están identificados (no son anónimos) y, por tanto, permiten análisis muy detallados. Por ejemplo, podremos discernir las valoraciones y apreciaciones por rango de calificación o por perfil inicial, etc.

Para respetar y fomentar la libre participación por parte de los alumnos en la evaluación del experimento, el compromiso al que llegamos era que el análisis de todos estos cuestionarios, junto con toda la información que almacena el sistema referente al seguimiento y compleción de las diferentes tareas no se procesará hasta que el curso haya finalizado. Esperamos que el estudio de estos datos permita generar una publicación con los resultados de la evaluación de la experiencia llevada a cabo durante el curso y promovida por el PIMCD.

4. Recursos humanos

El grupo de personas que ha llevado a cabo el proyecto es el siguiente: Luis Llana Díaz, Cristóbal Pareja Flores, Antonio Sarasa Cabezuelo, Jorge Carmona Ruber, Adrián Riesco Rodríguez y Carlos Gregorio Rodríguez. Todos ellos profesores del Departamento de Sistemas Informáticos y Computación.

Todos los integrantes tienen una larga trayectoria en innovación educativa que se refleja en forma objetiva en dos puntos concretos: la participación continua en proyectos de innovación educativa, especialmente en aquellos promovidos por la UCM, y las publicaciones de carácter docente.

La lista de proyectos de innovación en los que han participado alguno de los miembros del equipo es la siguiente: PIE 99/32, PIE 2001/28, PIE 2002/02, PIMCD 2005/34, PIMCD 2006/67, PIMCD 2008/33, PIMCD 2009/185, PIMCD 2009/160, PIMCD 2010/152, PIMCD 2010/248, PIMCD 2011/174, PIMCD 2011/334, PIMCD 2011/150, PIMCD 2013/220. El responsable del actual proyecto ya ocupó esa función con anterioridad en los proyectos PIE 2001/28 y PIMCD 2010/248.

Lista de publicaciones docentes de los miembros del equipo:

- Carlos Gregorio Rodríguez. Ergotracer: An Internet User Behaviour Tracer. Lecture Notes in Computer Science. 2346, pp. 121 - 132. 2002. ISSN 0302-9743
- Carlos Gregorio Rodríguez; Raquel Martínez Unanue; Ángel Herranz Nieva. Computing Curricula 2001. Novatica. pp. 47 - 54. 2002. ISSN 0211-2124
- Carlos Gregorio Rodríguez, Luis Llana Díaz, Raquel Martínez Unanue, Pedro Palao Gostanza, Cristóbal Pareja Flores y Ángel Velázquez Iturbide. EXercita: Automatic Publishing of Programming Exercises. ACM SIGCSE/SIGCUE International Conference on Innovation and Technology into Computer Science Education, ITiCSE'01. ACM Press. ISBN: 1-58113-330-8
- Carlos Gregorio Rodríguez, Luis Llana Díaz, Raquel Martínez Unanue, Pedro Palao Gostanza, Cristóbal Pareja Flores y Ángel Velázquez Iturbide. EXercita: A System for Archiving and Publishing Programming Exercises. Computers and Education, Towards an Interconnected Society, p.187-198, Kluwer, 2001. ISBN 0-7923-7188-7.
- Luis Llana, Enrique Martín, Cristóbal Pareja-Flores: FLOP, a Free Laboratory Of Programming. KOLI Calling 2012.
- Manuel Rubio-Sánchez, Päivi Kinnunen, Cristóbal Pareja-Flores, J. Ángel Velázquez-Iturbide: Lessons learned from using the automated assessment tool "Mooshak". Proceedings of SIIE'12: 209-214
- Alaeddin M. H. Alawawdeh, César Andrés, Luis Llana: CLAT: Collaborative Learning Adaptive Tutor. SEKE 2012: 747-752
- J. Ángel Velázquez-Iturbide, Cristóbal Pareja-Flores, Jaime Urquiza-Fuentes: An approach to effortless construction of program animations. Computers & Education (CE) 50(1):179-192 (2008)
- I. Hernán, C. Pareja, J. Ángel Velázquez: Testing-Based Automatic Grading: A Proposal from Bloom's Taxonomy. ICALT 2008: 847-849
- Cristóbal Pareja, J. Ángel Velázquez-I: Program Execution and Visualization on the Web. Encyclopedia of Information Science and Technology (IV) 2005: 23
- Cristóbal Pareja-Flores, J. Ángel Velázquez-Iturbide: Local versus comprehensive assignments: two complementary approaches. SIGCSE Bulletin 32(4): 48-51 (2000)

- Isabel Pita y Adrián Riesco. A Tool for Testing Data Type Implementations from Maude Algebraic Specifications. PROLE 2011. Electronic Notes in Theoretical Computer Science 282, páginas 61-71. Elsevier, 2012.
- M. Clavel y A. Riesco, The Web ITP Tool: A Quick Overview, en las actas del Segundo Congreso Internacional de Herramientas para la Enseñanza de la Lógica, Salamanca, Septiembre 2006.
- A Sarasa Un modelo de virtualización de los estudios superiores como marco para desarrollo de la convergencia europea en la educación superior Virtual Educa,2010
- A Sarasa, JM Dodero, ¿Cómo complementar la formación mediante proyectos de generación de contenidos web para una asignatura? Revista Latinoamericana de Tecnología Educativa-RELATEC 3 (1), 2007
- A Sarasa, J Miguel Cleva, S Estévez, N Morón, ¿Debemos superar WebCT?: hacia un modelo basado en objetos de aprendizaje Editorial Complutense ,2007
- A Sarasa, Usando la Wikipedia como motivación en el proceso de aprendizaje. RELATEC: Revista Latinoamericana de Tecnología Educativa 5 (2), 433-442, 2006
- A Sarasa, S Estévez, FJ Crespo, Creación de un curso sobre Metodología y Tecnología de la Información para su uso en un Campus Virtual Editorial Complutense, 2005
- A Sarasa, JM Dodero, ¿Son útiles los campus virtuales en la enseñanza superior?, JUTE 2005

Junto con el equipo del proyecto y de forma puntual, ha colaborado con nosotros una programadora pagada, como se especificaba en la petición del proyecto, con el dinero solicitado. Debido a la escasa financiación, no fue posible contar con sus servicios nada más que un periodo de tiempo muy pequeño (4 semanas). Este tiempo es muy insuficiente para llevar a cabo el desarrollo software de una herramienta, por tanto, la mayor parte del trabajo de programación, además del diseño, prueba y depuración del mismo, ha recaído sobre los miembros del equipo.

5. Desarrollo de las actividades

A grandes líneas, el desarrollo del trabajo estaba previsto en 3 fases, que tal y como figuraban en la propuesta de proyecto eran estas:

Fase 1: Análisis del estado del arte sobre la evaluación en MOOCs. En esta fase, además de la experiencia que tenemos en este ámbito, estudiaremos en más profundidad y reflexionaremos sobre los diferentes sistemas de evaluación usados en los MOOCs, tanto desde el punto de vista metodológico como técnico, pues estamos interesados no solamente en su uso sino también en su adaptación a las herramientas de las que disponemos, como por ejemplo Moodle.

Fase 2: Diseño, implementación e integración de las herramientas software necesarias para realizar diversas modalidades de evaluación para MOOCs. Construcción de los módulos de Moodle adecuados para soportar estas herramientas.

Fase 3: Experimentación de las metodologías de evaluación en un grupo grande. Deseamos poner en práctica diversas técnicas de evaluación de interés para los MOOCs, entre ellas habrá, al menos, evaluación masiva, evaluación por pares y autoevaluación. Utilizaremos para ello los alumnos matriculados en la asignatura de Informática obligatoria en primer curso en los grados que oferta la Facultad de Matemáticas.

Y además, ya adelantábamos que sería necesaria una última fase posterior a la terminación del proyecto dedicada a la publicación y difusión de los resultados y las experiencias obtenidas en congresos y revistas especializadas.

Si bien todas las fases se han cumplido, las mismas se han dilatado en el tiempo mucho más de lo inicialmente previsto. Por un lado, por la implicación del equipo y por la profundidad con la

que se han abordado muchos de los aspectos, por otro, por la mayor dificultad y carga de trabajo de los desarrollos e implementaciones de las herramientas, y, por último, pero muy importante, por la gran dedicación docente que tiene el correcto diseño de las nuevas actividades propuestas. Por ejemplo, para una actividad de evaluación por pares, es necesario: Diseñar cuidadosamente una rúbrica en la que aparezcan claramente los aspectos a evaluar. Cada aspecto a evaluar tiene una escala y cada elemento de la escala necesita una explicación detallada para evitar en lo posible ambigüedades en la interpretación. Antes de evaluar trabajos *reales* de sus compañeros, los alumnos evalúan unos trabajos de *prueba* cuya evaluación comparan con la dada por el profesorado. Estos trabajos de prueba tienen que estar muy bien pensados para que recojan mucha de la casuística posible y hagan hincapié en los aspectos que pueden ser más difíciles de evaluar.

A continuación describimos la duración aproximada y los trabajos ejecutados en cada una de las fases propuestas en el proyecto.

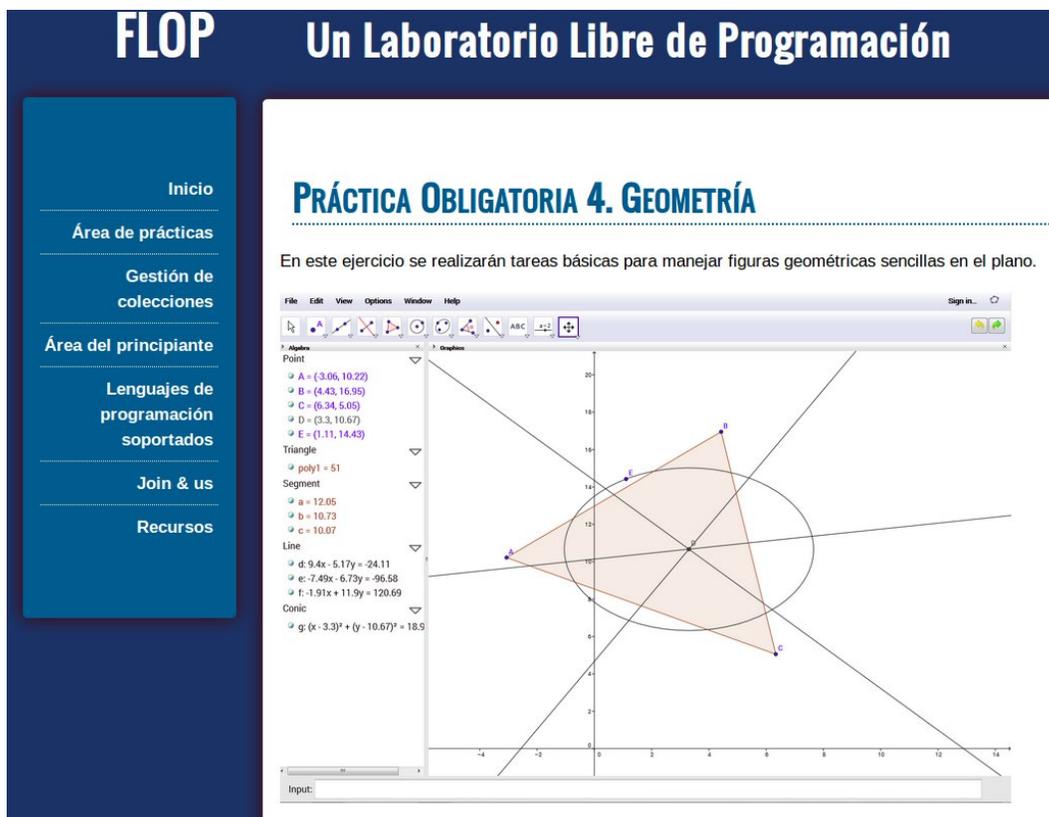


Figura 5: Detalle de la herramienta de corrección automática FLOP

Fase 1: Mayo-Julio 2014. En estas primeras semanas se hizo un estudio y un análisis de técnicas, metodologías y herramientas utilizadas en la evaluación de los MOOCs, para ello, aparte de la experiencia personal como participantes en MOOCs, se hizo un estudio en las principales plataformas Coursera, EdX, Udacity y MiriadaX.

En esta etapa, también recopilamos bibliografía especializada (bastante escasa) sobre el tema:

- Joanna Smith, Joe Tessler, Elliot Kramer, and Calvin Lin. Using Peer Review to Teach Software Testing. In Proceedings of the 9th annual international conference on International computing education research, ICER 2012, pages 93-98. ACM 2012.

- Jacob Lowell Bishop and Matthew A Verleger. The Flipped Classroom: A Survey of the Research. In Proceedings of the 120th ASEE Annual Conference & Exposition, June, 2013.
- Ihantola, P., Ahoniemi, T., Karavirta, V., & Seppälä, O. (2010). Review of recent systems for automatic assessment of programming assignments. In Proceedings of the 10th Koli calling international conference on computing education research. Koli Calling '10 (pp. 86–93). New York, NY, USA: ACM.
- Douce, C., Livingstone, D., & Orwell, J. (2005). Automatic test-based assessment of programming: A review. Journal of Educational Resources in Computing, 5(3) (article 4).
- Sheard, J., Dick, M., Markham, S., Macdonald, I., & Walsh, M. (2002). Cheating and plagiarism: Perceptions and practices of first year it students. SIGCSE Bulletin, 34, 183–187.

FLOP

- Inicio
- Área de prácticas
- Gestión de colecciones
- Área del principiante
- Lenguajes de programación soportados
- Join & us
- Recursos

Caso de prueba [T04_Segment]:
 Tiempo de CPU usado: 0.28001699999999996 seconds,
 Memoria usada: 21488 Kbytes.

Respuesta incorrecta

Failed test case. Input:

```
s = Segment(Point(0,0), Point(1,0))
print s, type(s),
s.distance(Point(2,0))
s.intersects(Point(0,0))
Segment(Point(4,4),Point(4,4))
s1,s2=Segment(Point(0,0),Point(2,0)),Segment(Point(3,0),Point(4,0))
print s1.intersects(s2),
print s1.distance(s2),
Segment(0,0,0,1)
Segment('a',0,0,0)
```

Expected output:

```
s = Segment(Point(0,0), Point(1,0)) : Ejecución correcta
print s, type(s), : Segment(Point(0,0),Point(1,0)) <class '__main__.Segment'> Ejecución correcta
s.distance(Point(2,0)) : Ejecución correcta
s.intersects(Point(0,0)) : Ejecución correcta
Segment(Point(4,4),Point(4,4)) : Excepción
s1,s2=Segment(Point(0,0),Point(2,0)),Segment(Point(3,0),Point(4,0)) : Ejecución correcta
print s1.intersects(s2), : False Ejecución correcta
print s1.distance(s2), : 1 Ejecución correcta
Segment(0,0,0,1) : Excepción
Segment('a',0,0,0) : Excepción
```

Figura 6: Detalle de una respuesta de la herramienta FLOP

Fase 2: Septiembre-Octubre 2014. El trabajo en estas semanas se dedicó a diseñar y definir las necesidades de las herramientas de software a utilizar para la evaluación; a estudiar el uso y la integración de herramientas existentes; y los requerimientos de los desarrollos de software a llevar a cabo para integrar en Moodle las actividades que permitían realizar este tipo de evaluaciones.

Utilizando un servidor propio de Moodle, que lleva años funcionando para las asignaturas de muchos de los profesores que forman parte del equipo, pudimos utilizar mucha de la funcionalidad más novedosa para permitir realizar un seguimiento detallado de la actividad del alumnado y diseñar las actividades de manera que se van ‘desplegando’ ante los participantes a medida que éstos avanzan por el curso (acceden a los documentos que se les facilita, realizan cuestionarios, entregan tareas, etc.)

En estas semanas también se diseñó de forma global el experimento en el que se englobarían el uso de las distintas técnicas de evaluación. Fue aquí cuando se decidió utilizar la metodología que se describe con detalle en el apartado 3.3. Puesto que el experimento se realizaría sobre una asignatura concreta y tendría repercusión en la evaluación de la misma, junto con el diseño del experimento, era también imprescindible realizar el diseño de la asignatura, no solo de contenidos, sino sobre todo de la forma adecuada de integrar las nuevas metodologías de evaluación por pares, autoevaluación y evaluación continua y, la forma en la que estas evaluaciones repercutirían en la evaluación final de la asignatura.

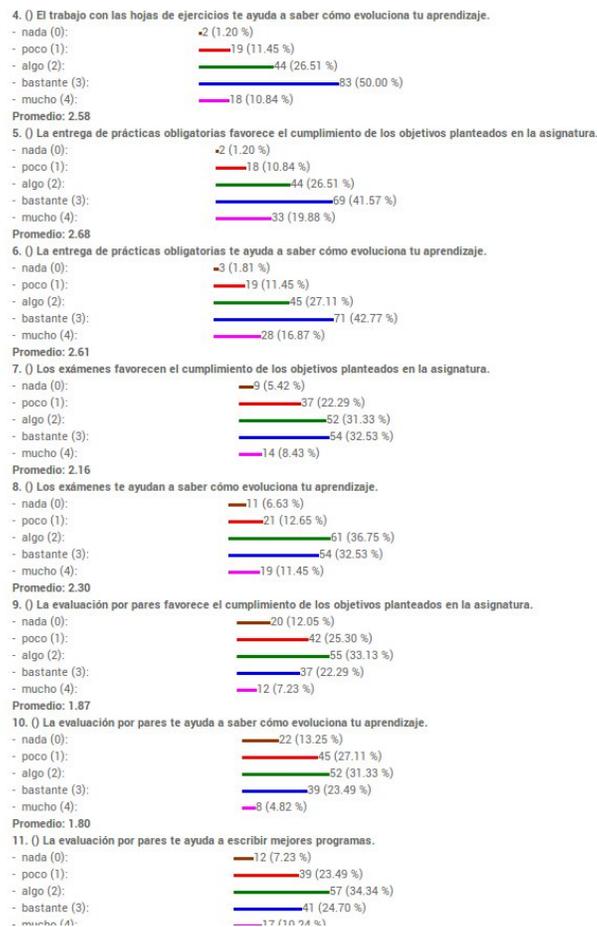


Figura 7: Algunos ejemplos de las estadísticas recopiladas

Fase 3: Noviembre 2014-Junio 2015. Esta fase se hizo un gran esfuerzo por integrar y poner todas las herramientas a funcionar. Dentro de esta fase, durante los últimos meses del 2014, se contrataron los servicios de un programador (de hecho una programadora) para que realizara las implementaciones de las herramientas diseñadas. Debido a la escasa financiación con la que cuentan estos proyectos, sólo se pudo contratar a una persona durante un mes, tiempo muy insuficiente para llevar a cabo un desarrollo de software.

El mayor peso en la implementación e integración de las herramientas ha recaído en los miembros del equipo, sobre todo de los profesores Luis Llana Díaz, Adrián Riesco Rodríguez, Jorge Carmona Ruber y Carlos Gregorio Rodríguez. Una breve descripción de las mismas es:

- Gestión y administración avanzada en Moodle para permitir el seguimiento y la particularización del campus virtual al avance personal de cada uno de los participantes.

- Desarrollo de un nuevo módulo de gestión de evaluaciones por pares para entregas de prácticas de programación. Este módulo gestiona las entregas, el reparto de trabajos a evaluar por cada alumno, las calificaciones obtenidas, etc. Además, también enlaza con un corrector automático FLOP (véanse las figuras 5 y 6) y con otra herramienta de evaluación de la calidad del código Pylint para obtener otras calificaciones de las tareas entregadas por los participantes. Con todos estos ingredientes, se pueden elaborar las calificaciones a la actividad de forma muy sofisticada.
- Diseño e integración de los tests de prueba que permiten automatizar la evaluación de la corrección de los trabajos de programación.
- Diseño e integración de las rúbricas de evaluación para cada una de las entregas. Diseño de ejemplos de rúbricas para ejercitar los criterios de evaluación antes de comenzar a evaluar.
- Diseño e integración de los formularios de autoevaluación con comportamiento dinámico.
- Búsqueda y utilización de una herramienta adecuada para la detección de plagios en las entregas de trabajos.
- Diseño y puesta en práctica de los cuestionarios y herramientas de evaluación del desarrollo de la experiencia por parte de los alumnos participantes.

Esta última fase se ha extendido durante todo el curso pues algunas de las tareas tenían que hacerse a medida de las actividades que se iban proponiendo, pero, sobre todo, porque a medida que avanzábamos en la asignatura y en el uso de las herramientas íbamos depurando errores y aumentando la funcionalidad y usabilidad.

Actualmente estamos desarrollando la última fase de estudio y análisis de resultados y preparación de publicaciones docentes en las que se muestre la experiencia y los resultados del proyecto de innovación llevado a cabo.

6. Conclusiones

El proyecto de innovación y mejora de la calidad docente que hemos desarrollado durante el curso 2014-2015 ha permitido utilizar técnicas novedosas de evaluación, en particular, evaluación automática y evaluación por pares, que pueden ser utilizadas en grupos grandes (MOOCs), aunque sin duda también son interesantes en grupos más reducidos.

Para ello, hemos tenido que trabajar y desarrollar tres aspectos fundamentales:

- Diseño de una asignatura que integrara las metodologías con las que queríamos experimentar autoevaluación, evaluación automática y evaluación por pares.
- Implementación e integración de las herramientas informáticas necesarias para poder llevarlas a cabo.
- Seguimiento del experimento para intentar medir las impresiones de los participantes y los resultados del mismo.

Actualmente, y una vez que el curso prácticamente ha concluido (aún falta la evaluación de la convocatoria de septiembre), estamos comenzando con el estudio y análisis de los resultados. Ésta es una tarea compleja, pues tenemos muchos datos que considerar: por un lado están los resultados obtenidos en las actividades planteadas a lo largo de todo el curso, por otro lado los resultados medidos de forma más clásica en exámenes tradicionales (necesarios por la obligada coordinación de todos los grupos de la asignatura), tenemos también todos los datos almacenados de forma pasiva por el servidor Moodle sobre la actividad de cada participante y, finalmente, también tenemos cuestionarios de evaluación realizados por los alumnos en distintas fases del experimento (véase por ejemplo la figura 7).

Aunque la evaluación de los resultados llevará aún un tiempo, lo que sí que podemos adelantar son algunas percepciones subjetivas de los integrantes del proyecto que además han sido profesores en los distintos grupos que han participado en el experimento.

- El campus virtual con seguimiento, particularizado a cada alumno y los cuestionarios de autoevaluación dinámicos favorecen el trabajo continuado de los estudiantes.
- La evaluación automática de ejercicios, en general, parece que está bien acogida entre los alumnos. Quizás porque permitíamos acceder a las herramientas de evaluación durante el desarrollo de los trabajos y así ellos mismos podían comprobar la corrección de sus prácticas.
- La evaluación por pares causa cierto rechazo inicial. La percepción de los alumnos es que supone un trabajo extra y que no aporta demasiado a su formación. La percepción del profesorado, sin embargo, es diferente: el rechazo es sobre todo porque no están acostumbrados a evaluar o valorar. El trabajo extra que los estudiantes tienen que realizar es indudable (hay que estudiar la rúbrica, hay que hacer las evaluaciones de prueba comparándolas con las dadas por el profesorado, finalmente hay que leer/estudiar el trabajo de los compañeros para proporcionar una calificación en cada uno de los aspectos indicados en la rúbrica. . .), pero ese trabajo desarrolla unas habilidades y fija unos estándares (los que indican las rúbricas) que luego utilizan en sus trabajos posteriores.

Para concluir nos gustaría también dejar algunas reflexiones con respecto al uso de estas técnicas en grupos muy grandes o en MOOCs. No podemos hablar de resultados definitivos, pues aún hay aspectos muy importantes a analizar en los datos recopilados, quizás el más importante es la adecuación de una evaluación por pares pura como método de calificación (en nuestro experimento, todas las prácticas han sido evaluadas por un profesor y así tenemos una evaluación *experta* con la que contrastar las evaluaciones de los participantes).

- Las técnicas de seguimiento y autoevaluación integradas en un campus virtual son fácilmente escalables a grupos muy grandes, si bien se necesita de un considerable trabajo inicial de diseño y desarrollo. Este último aspecto, hace que en grupos no muy grandes, este esfuerzo no sea muy rentable para el profesorado. También, las herramientas y técnicas pueden mejorar bastante para hacer que esta tarea sea más sencilla.
- La evaluación automática es una técnica que puede ahorrar mucho trabajo al profesorado siempre que los aspectos evaluables puedan ser objetivamente medibles. Esto claramente es así en el caso de las prácticas de programación para la asignatura de Informática: la evaluación se realiza mediante unas pruebas que el programa debe pasar. Los trabajos a entregar y las pruebas a pasar deben estar cuidadosamente diseñados pues la casuística de un programa de tamaño medio puede ser muy elevada, al igual que puede ser muy variada la forma de resolver problemas de los diferentes participantes. Igual que en el caso anterior, se necesita una masa crítica de estudiantes para que el esfuerzo de diseño de todos los componentes del sistema valga la pena.
- Para la detección de plagio (al menos en ejercicios de programación) existen herramientas que están funcionando muy bien, si bien su uso requiere de cierto entrenamiento y supervisión por parte del profesorado. Sin ayuda de una herramienta informática, esta tarea, por su complejidad y extensión, es difícilmente asumible para el profesorado incluso en grupos no muy grandes.
- La evaluación por pares une dos aspectos muy interesantes: por un lado es una actividad didáctica que fomenta importantes habilidades en los participantes, por otro, puede utilizarse para evaluar trabajos en grandes grupos.

Consideramos que la primera de las vertientes, la didáctica, es de indudable valor, si bien hay muchos aspectos interesantes a tener en cuenta.

- El estudiante tiene que familiarizarse con la técnica y empezar a utilizarla con ejercicios sencillos y muy guiados.
- La cantidad de trabajos que un participante puede revisar es inversamente proporcional al interés de estos trabajos. El trabajo de revisión es productivo para el revisor si el trabajo a revisar es interesante (y éste es un concepto escurridizo). Como el número de revisiones para trabajos complejos e interesantes es reducido, hay que tener especial atención en la forma de repartir los trabajos para que a cada participante le sean asignados trabajos a revisar que puedan proporcionar un desarrollo de sus habilidades.
- El adecuado diseño de la actividad: trabajo a entregar, rúbrica, ejemplos, etc. supone un gran trabajo para el docente.
- Las herramientas de gestión automática permiten escalar bastante bien la tarea de entrega, reparto, asignación de las evaluaciones, etc.

Con respecto al uso de la evaluación por pares como herramienta de evaluación completamente independiente (sin intervención directa del profesorado), nos gustaría hacer las siguientes consideraciones:

- Se necesita un diseño muy cuidadoso de la actividad, además de los aspectos comentados anteriormente que forman parte intrínseca de la actividad, (como la rúbrica, ejemplos. . .), cada uno de los aspectos en la rúbrica tiene que ser ponderado de cara a una calificación final. Este aspecto requiere de cierta práctica pues en ocasiones, más de las deseadas, hay entregas de trabajos *patológicos* que pueden tener una calificación aceptable en la rúbrica y que con un método de evaluación más global no la tendrían.
- Como comentábamos en el apartado didáctico, si los trabajos son interesantes y, por tanto, tienen una cierta extensión/complejidad, no se puede pedir a los participantes que evalúen un gran número de ellos. El que un trabajo tenga pocas evaluaciones hace que la calificación obtenida por el mismo sea muy sensible a las variaciones en las formas de evaluar de los diferentes participantes. Este es un tema que requiere de un buen estudio formal para encontrar medidas de *fiabilidad* en estas evaluaciones de los pares.
- Consideramos que un buen complemento a la evaluación por pares, siempre que pueda llevarse a cabo, es la evaluación automática. De esta forma, hay una evaluación objetiva con la que contrastar las evaluaciones recibidas.

Todas estas consideraciones seguirán siendo objeto de estudio pues para el curso 2015-2016 seguiremos completando y desarrollando la experiencia en la asignatura de Informática en los primeros cursos de los grados ofertados en la Facultad de Matemáticas. Esperamos también reflejar todo este trabajo y los resultados obtenidos del mismo en publicaciones en revistas y congresos especializados.