

Recomendaciones formativas para la evaluación sostenible de proyectos colaborativos en forjas de software

Juan Manuel Dodero, Iván Ruiz Rube, Ignacio Traverso Ribón, Manuel Palomo Duarte

Departamento de Ingeniería Informática, Escuela Superior de Ingeniería

juanma.dodero@uca.es

ABSTRACT: La evaluación orientada al aprendizaje pretende medir y registrar el conocimiento, las habilidades y las actitudes de los estudiantes con un propósito formativo. En las experiencias de aprendizaje basadas en proyectos, el seguimiento detallado de las actividades en que participan los miembros de cada equipo pueden ser beneficiosas para la evaluación de la labor de los estudiantes. Los profesores pueden evaluar las habilidades de trabajo en equipo mediante el seguimiento de un procedimiento de evaluación y la compleción de un conjunto de instrumentos de evaluación, como las rúbricas, aplicándolas a cada equipo y cada miembro. Con el objetivo de hacer de la evaluación una labor sostenible y formativa para los estudiantes se suelen emplear estrategias como la auto-evaluación, la evaluación por pares y la co-evaluación. La conducción de tales estrategias colaborativas no es fácil para los estudiantes, dado que, antes de ser capaces de realizar evaluaciones con un cierto valor formativo, necesitan tener una comprensión razonable del procedimiento de evaluación y de sus instrumentos, así como una realimentación sobre el porqué de las evaluaciones. Este trabajo pretende elaborar un método que ayude a proporcionar recomendaciones de evaluación en estrategias colaborativas, de forma que los evaluadores puedan obtener una realimentación formativa al realizarlas. Con tal motivo se ha elaborado una rúbrica de evaluación de proyectos de ingeniería de software y se han seleccionado una serie de indicadores que puedan proporcionar evidencias sobre las habilidades desarrolladas durante el trabajo. El instrumento fue aplicado en las prácticas de la asignatura de Ingeniería Web, consistente en el desarrollo de un proyecto a escala para crear una aplicación web. La gestión de cada proyecto se realizó en una forja de software, consistente en una herramienta de *ticketing* para la planificación y seguimiento de tareas, un repositorio de control de versiones para depositar los entregables de software y un wiki para alojar los entregables de texto. Las métricas usuales disponibles en las forjas de software se ampliaron con herramientas adicionales de medida de la calidad del código fuente y de las contribuciones realizadas al wiki. La experiencia proporciona evidencias a favor de construir recomendaciones de evaluación de ciertas competencias, de forma que se facilite la labor a evaluadores noveles como los estudiantes, que deben completar un procedimiento de evaluación formativa, así como hacer los procesos de evaluación más sostenibles.

KEY WORDS: proyecto, innovación, mejora, docente, aprendizaje basado en proyectos, evaluación colaborativa, forjas de software

INTRODUCCIÓN

Una evaluación sostenible del aprendizaje persigue desarrollar en los estudiantes capacidades de evaluación, basándose en estrategias de evaluación colaborativa, auto-evaluación y evaluación entre iguales [2]. Para ello los estudiantes utilizan instrumentos de evaluación prediseñados que permitan evaluar los resultados de su aprendizaje [1]. Las habilidades de evaluación se adquieren elaborando juicios y proporcionando explicaciones para esos juicios en forma de realimentación [10]. Esto se ha probado de manera favorable en evaluaciones formativas que incluían realimentación y auto-supervisión explícitas como parte de las tareas de los estudiantes [3]. Por otro lado, los instructores quieren ser conscientes de no obviar información relevante para la evaluación sumativa [5]. Una evaluación sostenible dota a los estudiantes para progresar con confianza en el aprendizaje a lo largo de la vida, sin por ello aumentar la carga del personal académico [7].

Motivación

Los procedimientos e instrumentos de evaluación más eficaces están basados en la inspección de los resultados del aprendizaje como entregables explícitos generados por los estudiantes. En entornos virtuales de aprendizaje (VLE, *Virtual Learning Environment*), los procedimientos de e-evaluación constituyen interacciones guiadas durante las cuales instructores y estudiantes deben inspeccionar un cierto

número de entregables y proporcionar sus evaluaciones rellenando formularios (i.e. instrumentos) estructurados con criterios de evaluación, tales como las rúbricas y las escalas de valoración [11]. Durante los procedimientos de auto-evaluación y evaluación entre iguales, resulta difícil para los estudiantes aprender a estimar valoraciones y clasificaciones, así como a proporcionar explicaciones sobre los motivos de sus juicios. En el lado del profesor, los procedimientos de evaluación no se adaptan bien ante un aumento del número de alumnos, grupos o resultados a evaluar. Aunque la carga del evaluador puede aliviarse haciendo que los estudiantes formen parte del procedimiento, no es deseable dejar su efectividad completamente en manos de los estudiantes, que no suelen estar acostumbrados a evaluar y pueden adolecer de la capacidad de auto-crítica.

El aprendizaje por proyectos

La evaluación de los resultados del aprendizaje es especialmente gravosa cuando se llevan a cabo estrategias complejas de aprendizaje. Por ejemplo, en experiencias de aprendizaje por proyectos (PBL, *Project-Based Learning*), cada proyecto conlleva un conjunto explícito de resultados en forma de múltiples productos que deben ser construidos por los estudiantes y que forman la base de planes de evaluación y rúbricas como instrumentos de evaluación [8]. El instructor de una experiencia de PBL puede gestionar con éxito el proceso de aprendizaje mediante el empleo de algunas herramientas

que proporcionan cierta sistematización y la posibilidad de medir los resultados intermedios del proceso. En este aspecto, las forjas de software como *assembla* (www.assembla.com) o *redmine* (www.redmine.org) ofrecen aplicaciones web para la gestión y seguimiento de proyectos, incluyendo sistemas de *ticketing* para controlar tareas planificadas e incidencias. También proporcionan sistemas de control de versiones (VCS, *Version Control System*) para depositar y compartir código fuente y herramientas de tipo wiki para colaborar en la escritura de entregables textuales. En el caso de los proyectos de software, los sistemas de revisión de código fuente como *sonar* (www.sonarsource.org) son muy útiles para supervisar la calidad de los entregables en cualquier momento del proceso de PBL.

PROBLEMÁTICA Y LIMITACIONES

Esta actuación pretende responder a *¿cómo pueden los profesores evaluar y los estudiantes aprender a evaluar los resultados de una experiencia compleja de aprendizaje de una manera sostenible?* Dar recomendaciones de evaluación puede ayudar a los estudiantes a la hora de elaborar la proalimentación necesaria para que la evaluación sea eficaz. La provisión de recomendaciones de evaluación puede hacer que el proceso de evaluación sea más sostenible (Boud, 2000) incluso desde el punto de vista de los profesores.

Como los procesos complejos de aprendizaje pueden ser muy variados, esta actuación se limita al desarrollo de una experiencia de PBL desplegada con la ayuda de una forja de software. A partir de los registros de actividad recogidas por la forja, se han determinado indicadores y elaborado pistas que proporcionen recomendaciones a los estudiantes sobre cómo evaluar ciertas competencias, de forma que los juicios críticos y las explicaciones emitidas por los estudiantes estén refrendados en evidencias recogidas por la forja. De esta forma, se espera que la evaluación de un proceso de PBL sea más eficaz y sostenible.

CONTEXTO Y METODOLOGÍA

La asignatura objeto de la actuación es *Ingeniería Web* (IW) de 5º curso de Ingeniería Informática [4]. Las competencias más significativas que se desarrollan en la asignatura son la capacidad de organización y planificación y el trabajo en equipo (transversales), el conocimiento de métodos, técnicas y herramientas de desarrollo de aplicaciones Web (cognitivas) y su aplicación al ciclo completo de ingeniería de una aplicación Web (procedimentales), además de la motivación por la calidad del desarrollo y el aprendizaje autónomo de nuevos métodos y herramientas (actitudinales) [6]. El método de aprendizaje es una combinación de aprendizaje colaborativo y basado en proyectos. Durante la primera parte del curso (7 semanas) desarrollan una serie de trabajos colaborativos basados en un wiki, enfocados al aprendizaje de nuevos métodos y entornos de desarrollo, mientras que en la segunda parte del curso (8 semanas) desarrollan en equipo una aplicación web a escala con todas las fases del ciclo de vida de la ingeniería del software.

Esta actuación de innovación docente, enfocada al desarrollo de capacidades de evaluación, fue puesta en práctica en la segunda parte de la asignatura con un enfoque de PBL. Para ello, en primer lugar se elaboró un instrumento

de evaluación de tipo rúbrica con las competencias que se deseaba evaluar en los estudiantes. Dicha rúbrica se difundió entre los estudiantes para familiarizarlos con las competencias a evaluar.

Durante las clases de prácticas se celebraron varias sesiones donde se ejemplificó a los estudiantes cómo evaluar las competencias de la rúbrica a partir de las métricas disponibles en la forja de software y en los instrumentos de análisis de los entregables. Como los criterios que evalúa dicha rúbrica son numerosos (hasta 32 criterios), y cada uno se puede evaluar a través de varios resultados del aprendizaje, las sesiones formativas de evaluación sostenible se centraron en los siguientes criterios, que permitían evaluar competencias tanto transversales como específicas:

- C1: el equilibrio en el reparto de tareas a los distintos roles del proyecto (competencias transversales instrumentales “capacidad de organización y planificación” y “capacidad para resolver problemas”; competencia transversal personal “trabajo en equipo”)
- C2: el uso de herramientas de las forjas de desarrollo de software (competencia específica cognitiva “conocer herramientas de desarrollo web”)
- C3: el cumplimiento de tareas e hitos en el calendario (competencia transversal instrumental “capacidad de organización y planificación”)

Los entregables de texto del proyecto se depositaron en un wiki, mientras que el código fuente se depositó en el sistema de control de versiones (VCS) de una forja software. Se propusieron ejemplos de evaluación sobre los aspectos anteriores. A continuación se solicitó a los estudiantes realizar evaluaciones formativas sobre los entregables del propio equipo y de los de otros equipos. Finalmente, el profesor realizó una evaluación final sumativa de los proyectos, que publicó y discutió en co-evaluación con los estudiantes.

DESARROLLO DE LA EXPERIENCIA DE EVALUACIÓN SOSTENIBLE

Para elaborar las recomendaciones de evaluación sobre la rúbrica se propusieron un conjunto de indicadores analizables a través de la forja, el wiki y las herramientas de apoyo.

Herramientas de apoyo

Las herramientas usadas para explorar las evidencias sobre el desarrollo de las competencias investigadas fueron las siguientes:

- *Assembla tickets* como sistema de planificación, gestión y seguimiento de tareas
- *Subversion* o *Git* (a elección de cada equipo) como repositorio VCS de código fuente
- *Mediawiki* y *StatMediaWiki* [9] para albergar y analizar los entregables textuales
- *Sonar Source* como herramienta de análisis de la calidad de los entregables de código fuente.

Indicadores

Los siguientes indicadores se emplearon para observar evidencias en cada uno de las herramientas de apoyo:

a) ASSEMBLA:

- Retraso de finalización del último ticket de cada milestone

- Número de tickets abiertos, asignados y resueltos por cada usuario y cada equipo
- Diferencia de tiempo entre la creación y el cierre de cada ticket
- Diferencia entre el número de tickets creados y cerrados/resueltos en un cierto periodo (*kill rate*)

b) StatMediaWiki:

- Número de ediciones de texto y cantidad de texto aportado
- Clasificación de cada individuo dentro del equipo como contribuidor al wiki
- Reparto en la distribución del número de contribuciones a lo largo de la vida del wiki

c) SVN/GIT:

- Número de aportes (commits) de cada individuo y total del equipo
- Diferencia de tiempo entre cada par de commits

d) SONAR Source:

- Índice de cumplimiento de reglas de SONAR
- Clasificación de cada individuo en el repositorio de código fuente

Analítica y discusión

Cada uno de los indicadores disponibles en las herramientas web anteriores se empleó como fuente de evidencias para los criterios a evaluar. Por ejemplo, el criterio C1 de "equilibrio en el reparto de tareas" se observó y analizó principalmente a partir de dos fuentes: el número de *tickets* (tareas) asignadas en *Assembla* y el número de ediciones de texto en el wiki de entregables. En las siguientes direcciones pueden observarse los informes de la distribución de tickets y ediciones entre los miembros del equipo de un proyecto.

- Tickets: <http://bit.ly/tckblnc>
- Wiki: <http://bit.ly/wkiblnc>

Los indicadores basados en tickets asignados y resueltos, usados para evaluar el criterio C1 a favor de la competencia de "capacidad de organización", pueden proporcionar evidencias a favor de otras competencia. En concreto, podrían servir para evaluar la "capacidad de resolución de problemas", dado que los tickets se emplean no solo para gestionar las tareas planificadas, sino también las emergentes (indidencias o *issues*). Estas tareas suelen ser creadas durante la ejecución del proyecto y pueden ser asignadas a personas que desempeñan un rol distinto del planificado inicialmente, en un comportamiento conocido como *backup behaviour*. Sin embargo, como las herramientas usadas no permiten definir formalmente los roles de cada miembro del equipo, no es sencillo obtener un indicador que proporcione evidencias de dicho comportamiento.

Aunque existen herramientas que permiten medir el nivel de contribuciones de código fuente de cada miembro del equipo, éstas no fueron empleadas y por tanto tenidas en cuenta en el análisis. Sin embargo, sí se analizó el número total de *commits* del equipo para medir el grado de uso de las herramientas de VCS. Dicha evaluación se acompañó de otro indicador que medía el número total de tareas creadas en *Assembla Tickets*. La combinación de ambos indicadores permitió evaluar el grado de uso de las herramientas de la forja de desarrollo de software (criterio C2) a nivel de equipo.

El criterio C3 sobre el cumplimiento de los hitos en el calendario se analizó a partir de la herramienta *Assembla Milestones*, que ofrece informes sobre el cumplimiento de las fechas fijadas en los hitos planificados y el esfuerzo (estimado vs. reportado) para alcanzar cada hito. En la dirección <http://bit.ly/mlstons> puede verse el informe de hitos de un proyecto. Otro indicador favorable al criterio C3 es la evolución del cumplimiento de las tareas de cada hito, medible a través del análisis del *kill rate* semanal, del puede observarse un ejemplo en <http://bit.ly/klrate>.

Resultados

Las figuras 1-3 proporcionan una visión comparativa de las calificaciones (de 1 a 10) recibidas por los alumnos en la evaluación sumativa de los últimos cursos académicos. El número de participantes en la experiencia fue de 7 equipos en todos los cursos académicos, exceptuando 2009 que hubo solo 4 equipos. El número de miembros de cada equipo era de 4 personas. La única diferencia del curso 2012 es que en él tuvo lugar la actuación de evaluación sostenible basada en recomendaciones descrita en este trabajo, y ausente en los semestres anteriore. La Figura 1 representa la evolución de la calificación del criterio C1 (reparto equilibrado de tareas), la Figura 2 para el criterio C2 (uso de herramientas de desarrollo) y la Figura 3 para el criterio C3 (cumplimiento de tareas e hitos).

De una observación exploratoria de la evolución de la calificaciones, se observan evidencias a favor de la actuación realizada en el semestre académico 2012. No obstante, ya en el semestre anterior se alcanzaron resultados similares, sin haber realizado las acciones formativas de evaluación colaborativa con los estudiantes. Cuando se dispongan de datos de más semestres y competencias, un análisis más formal y estadísticamente más significativo podría determinar la influencia de estas prácticas de evaluación en el resultado final de las competencias evaluadas en los estudiantes. Por el momento, los resultados señalan que no parece haber evidencias significativas en contra de las prácticas de evaluación sostenible de cara a la evaluación final de los estudiantes.

CONCLUSIONES

La experiencia desarrollada proporcionó indicios a favor de la realización de prácticas de evaluación sostenible en una experiencia de PBL, basada en la formación de los estudiantes en un conjunto de indicadores que les proporcionen recomendaciones de evaluación. Como trabajo futuro, se plantea integrar dichas recomendaciones de evaluación en forma de proalimentación en las rúbricas empleadas para la evaluación formativa.

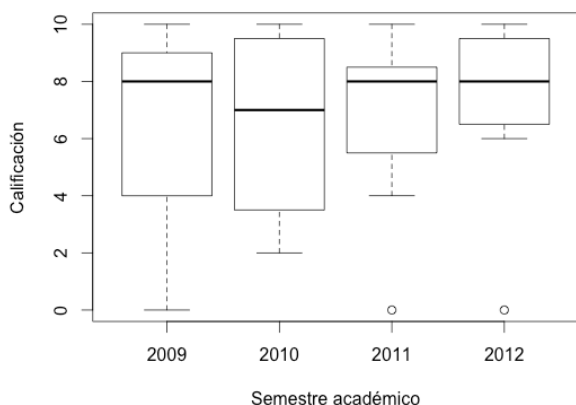


Figura 1. Evolución de la calificación de la evaluación sumativa del criterio C1: reparto equilibrado de tareas.

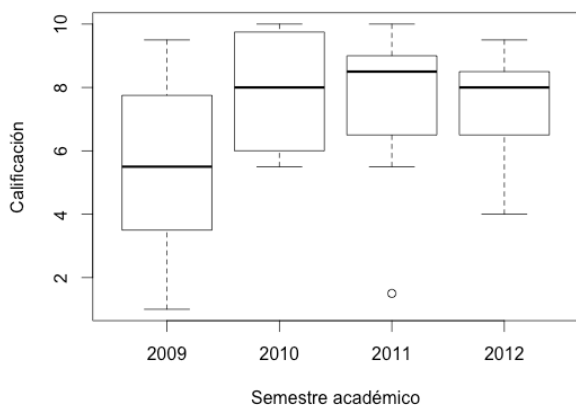


Figura 2. Evolución de la calificación de la evaluación sumativa del criterio C2: uso de herramientas de desarrollo.

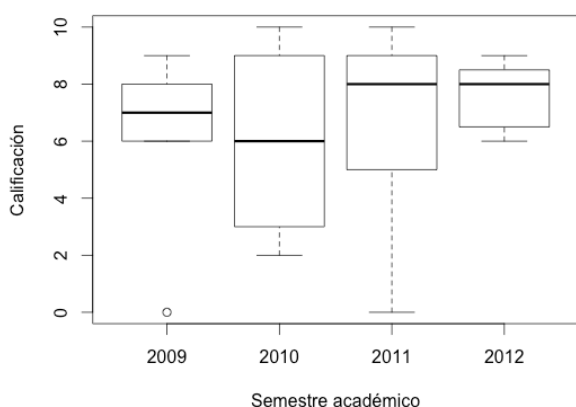


Figura 3. Evolución de la calificación de la evaluación sumativa del criterio C3: cumplimiento de tareas e hitos.

REFERENCIAS

1. Barootchi, N. & Keshavarz, M. Assessment of achievement through portfolios and teacher-made tests. *Educational Research*, 44 (3), pp. 279–287, 2002.
2. Boud, D. Sustainable Assessment: Rethinking assessment for the learning society, *Studies in Continuing Education*, 22(2), pp. 151-167, 2000.
3. Boud, D. Reframing assessment as if learning were important. En Boud, D. and Falchikov, N. (eds.), *Rethinking Assessment in Higher Education. Learning for the longer term*, Routledge, London, pp. 14–25, 2007.
4. Dodero, J. M. , Wiki de la asignatura Ingeniería Web. Disponible en http://deki.uca.es/Docencia/IW_1713056, Universidad de Cádiz, 2012.
5. Falchikov, N. *Improving Assessment Through Student Involvement. Practical solutions for aiding learning in higher and further education*. RoutledgeFalmer, London, 2005.
6. Hernández-Leo, D., Moreno, V., Dodero, J.M., Pardo, A., Romero-Ternerero, M. C., Dimitriadis, Y., Asensio-Pérez, J. I. Aplicación de Recomendaciones para la Alineación de Competencias, Metodología y Evaluación en Asignaturas de Ingeniería Telemática, Informática y Electrónica. *Revista Iberoamericana de Tecnologías del Aprendizaje*, 7(1), pp. 13-20, 2012.
7. JISC, Effective Assessment in a Digital Age. A guide to technology-enhanced assessment and feedback, Tech. Rep. Higher Education Funding Council for England, 2010.
8. Markham T, Larmer J, Ravitz J. *Project Based Learning Handbook: A Guide to Standards-Focused Project Based Learning for Middle and High School Teachers*, 2nd ed. Oakland, CA, Buck Inst for Education; 2003.
9. Palomo-Duarte, M., Dodero, J.M. Medina-Bulo, I., Rodríguez-Posada, E. J., Ruiz-Rube, I. Assessment of collaborative learning experiences by graphical analysis of wiki contributions. *Interactive Learning Environments*, [to be published], pp. 1-23, 2012.
10. Sadler D., Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*. 35(5), pp. 535-550, 2009.
11. Walvoord, B.E., Anderson, V.J. *Effective grading: A tool for learning and assessment in college*, 2nd ed. San Francisco, CA, Jossey-Bass, 2009.

ANEXOS

Algunos ejemplos de informes de los indicadores utilizados por los estudiantes para la auto-evaluación y evaluación entre iguales pueden hallarse en <http://bitly.com/bundles/dodero/1>