

PASSAM: Peer ASSESSMENT And Monitoring system

Juan A. Marin-Garcia^a, Jose P. Garcia-Sabater^b, Joan Morant Llorca^c, J. Alberto Conejero^d

^a ROGLE. Dpto. de Organización de Empresas. Universitat Politècnica de València. Camino de Vera S/N 46021 Valencia. jamarin@omp.upv.es, ^b ROGLE. Dpto. de Organización de Empresas. Universitat Politècnica de València. Camino de Vera S/N 46021 Valencia. jpgarcia@omp.upv.es, ^c ROGLE. Dpto. de Organización de Empresas. Universitat Politècnica de València. Camino de Vera S/N 46021 Valencia. jmorantllorca@gmail.com ^d ETS Ingeniería Informática. Dpto. de Matemática. Universitat Politècnica de València. Camino de Vera S/N 46021 Valencia. aconejero@upv.es.

Abstract

We present a tool developed by the authors to facilitate the evaluation process by pairwise comparison and apply it in a context of problem-based learning to encourage reflection and previous work of students. Experience has proven successful with small groups (20 and 14 students) of undergraduate and master students. It remains for future research to expand the experience to large groups and check the psychometric properties of the marks obtained with this method.

Keywords: *peer assessment, tutor assessment, paired comparison, motivation, homework, continuous assessment.*

Resumen

Presentamos una herramienta desarrollada por los autores para facilitar el proceso de evaluación por comparación pareada y la aplicamos en un contexto de aprendizaje basado en problemas para favorecer la reflexión y trabajo previo de los alumnos. La experiencia ha resultado exitosa con grupos pequeños (20 y 14 alumnos) de grado y máster. Queda pendiente para investigación futura ampliar la experiencia a grupos numerosos y comprobar las propiedades psicométricas de las puntuaciones obtenidas con este método.

Palabras clave: *evaluación por pares, evaluación del profesor, comparación pareada, motivación, deberes, evaluación continua.*

1. Introducción

La evaluación de trabajos, actividades o competencias de los alumnos es una actividad que suele consumir mucho tiempo y no está exenta de sesgos (lenidad, severidad, halo, recencia, primera impresión....) (Marin-Garcia et al., 2012, Dolan et al., 2007). Estos sesgos pueden generar dudas, en la persona que evalúa, acerca de si ha evaluado a los alumnos de manera coherente. Es decir, si el orden de las puntuaciones refleja el grado de superación de los objetivos de aprendizaje (Marin-Garcia et al., 2015a, Marin-Garcia et al., 2015b). Estos problemas son más patentes cuando el número de alumnos aumenta (por ejemplo, por encima de 30-40 alumnos).

Una de las opciones para mejorar la calidad de las evaluaciones es usar la comparación pareada. En ella, en lugar de evaluar un trabajo independientemente frente a una escala numérica o una descripción verbal, se comparan dos objetos y se valora cuál de ellos cumple mejor los criterios establecidos para la comparación (Marin-Garcia et al., 2015b).

Por otra parte, consideramos que para garantizar un correcto aprendizaje usando un caso/problema a resolver en grupo, durante una sesión presencial de clase, es necesario que los alumnos hayan trabajado el caso antes de la clase, de otro modo no se produce una reflexión y maduración del problema y la eficacia de la metodología se resiente (Marin-Garcia et al., 2008). En este sentido usar a los alumnos como evaluadores de trabajos de sus compañeros permite implicarles en las actividades previas a las sesiones.

2. Objetivos

El objetivo de este trabajo es comprobar si la herramienta desarrollada por los autores, PASSAM (Peer ASSESSMENT And Monitoring system), permite realizar evaluaciones de manera rápida y cómoda, usando el método de comparación pareada. Para ello usaremos tanto al profesor y a los alumnos como evaluadores.

En esta experiencia unimos varias actividades relacionadas con las metodologías activas (Marin-Garcia and Conchado-Peiró, 2012, Lloret and Marin-Garcia, 2008, Marin-Garcia, 2007, de la Hoz i Casas and de Blas del Hoyo, 2009, Componation and Farrington, 2000). Por un lado, presentamos una actividad asociada a la “Flip Teaching” en la que los alumnos deben realizar una reflexión individual como preparación previa al trabajo en grupo en el aula (Marin-Garcia et al., 2008) en un contexto de Aprendizaje Basado en Problemas (Woods, 1996, Hansen, 2003, Northwood et al., 2003, Gibbings and Brodie, 2008, Spronken-Smith and Harland, 2009, Downing et al., 2009). Por otro, incorporamos la evaluación por pares (Stefani, 1994, Hanrahan and Isaacs, 2001, Prins et al., 2005, Zhang et al., 2008, Marin-Garcia, 2009) como método para que los alumnos estén al tanto del trabajo de sus compañeros (reflexión crítica antes del trabajo de grupo) y para proporcionar un feedback rápido a los alumnos, de modo que sirva de filtro y motivación. Para ello, nos ayudamos de una herramienta web (PASSAM), desarrollada por los autores para facilitar el proceso de evaluación y el cálculo de las puntuaciones.

3. Desarrollo de la innovación

Hemos realizado la experiencia en dos asignaturas con un solo grupo y con pocos alumnos. Una asignatura de grado con 20 alumnos y una asignatura de máster con 14.

En cursos anteriores hemos constatado que pocos alumnos leían y reflexionaban el enunciado del caso antes de la clase en la que se tenía que resolver. Para fomentar que hagan “los deberes”, se les ha pedido a los alumnos que dibujen su interpretación del caso antes de asistir a la siguiente clase. Los dibujos tienen la ventaja de que representan diferentes visiones de un texto mediante un lenguaje que no es el texto, por tanto se ve obligado a reinterpretar el caso. Algunos alumnos tienen una gran resistencia dibujar (y mucho más a mano) y, a lo largo del curso se les anima a que lo hagan porque no es la calidad del dibujo lo que importa, sino la reflexión que hacen para poder dibujarlo. La única

restricción es que el dibujo sea personal. También se les enseña a mejorar sus dibujos prestando atención a los contenidos (tanto hechos como percepciones) y aspectos formales que facilitan la comunicación. Además, se les anima a que realicen dibujos ricos en detalles (al menos tan complejos como el problema que quieren resolver). Un dibujo con pocos elementos (por bien dibujado que este) suele indicar poca profundidad en el análisis.

En las primeras versiones de aplicación de la idea, el dibujo se subía a una presentación de “Google Drive” compartida con acceso de escritura. Pero con alumnos de grado, todos los dibujos eran muy parecidos al primero (que se tomaba como ejemplo para los siguientes). Con lo que redujimos el esfuerzo a realizar por el alumno y empeoró la calidad de lo observado. Con alumnos de Master no hemos detectado ese problema de momento pues los alumnos respetan el orden sugerido: primero dibujan, luego hacen una foto del dibujo y solo acceden a la presentación compartida cuando van a pegar la foto en ella.

En una segunda experiencia, se subían los dibujos al “Learning Management System” (LMS) Sakai/poliformat (Steel, 2009, Rubin et al., 2010) y el profesor tenía que descargarlos uno a uno, montarlos en una presentación y luego analizarlos en clase en función de la idea del profesor. Esto representaba mucho trabajo. Pero todos los alumnos querían ver su dibujo en la pantalla. Mientras la participación era escasa (o el grupo tenía menos de 15 alumnos), no representaba un problema excesivo. Sin embargo, con grupos grandes es inviable (por las horas invertidas por el profesor y por el tiempo de clase que se consume en presentar los dibujos).

En la versión actual, los alumnos suben sus dibujos a Sakai/Poliformat, el profesor se los descarga en un archivo zip que genera automáticamente el LMS. Hemos creado una aplicación web que importa ese Zip y genera las parejas de dibujos a comparar. Los alumnos pueden acceder a la aplicación con su móvil (les pasamos una URL corta y un código QR) y votar. La aplicación les presenta parejas de dibujos y el alumno debe elegir el que más les gustaría que fuese discutido en clase (o cualquier otro criterio que quiera establecer el profesor). Los alumnos no comparan las mismas parejas de dibujos (la aplicación creada tiene un algoritmo que intenta que todos los dibujos se vean por igual y optimiza las comparaciones para poder crear una lista ordenada desde el más preferido a menos preferido). El profesor también puede participar en las votaciones (de momento con igual peso que un alumno, pero eso se puede configurar en próximas versiones).

En la clase presencial, antes de que los alumnos empiecen a resolver el caso, el profesor utiliza los dibujos más votados (o los perdedores), para revisar los principales aspectos del caso, comentar cada parte con detalle y utilizamos el dibujo para hablar y fijar el problema. Gracias al filtro realizado por los propios alumnos, no importa que no se vean todos los dibujos en clase. Los alumnos saben que todas las participaciones han sido valiosas porque todos sus compañeros han podido ver su dibujo. Gracias al filtro realizado, la fase de revisión de los dibujos se mantiene en torno a los 15-20 minutos planificados para esta fase de la actividad.

4. Resultados

En la implantación de este año hemos conseguido:

- a) Que la mayoría de los alumnos realicen la tarea previa y las evaluaciones de sus compañeros
- b) La comprensión del caso y las soluciones presentadas en clase han sido de mayor calidad que años anteriores (nosotros lo atribuimos a una mayor maduración y reflexión del caso por parte de los alumnos)
- c) El profesor puede revisar rápidamente en clase los trabajos más votados por los alumnos (los alumnos aceptan mejor este “filtrado” que el que pudiera hacer el profesor, pues ellos han participado en el proceso).

A continuación (Figura 1) podemos ver algunos dibujos de ejemplo realizados por los alumnos:

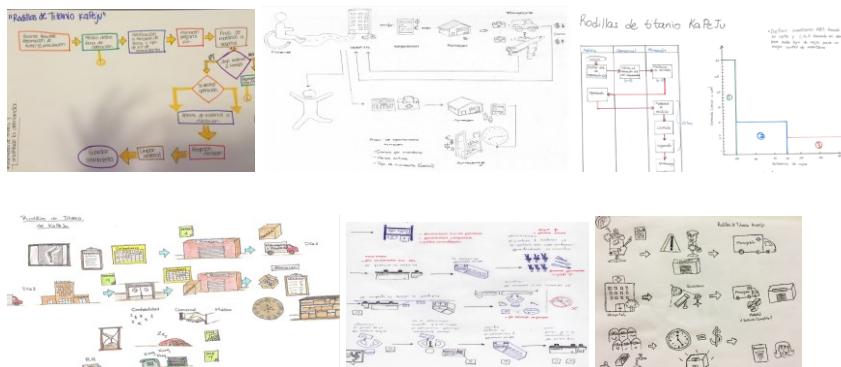


Figura 1. Ejemplos de dibujos de los alumnos

Y en la (Figura 2) mostramos la pantalla de comparación que usan los alumnos.

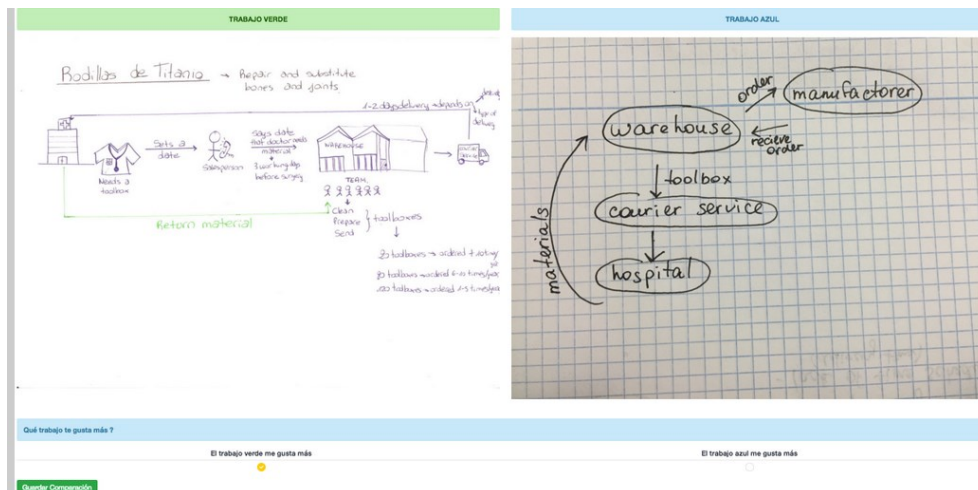


Figura 2. Pantalla de comparación a la que acceden los alumnos siguiendo el código QR

Por otra parte, hemos hecho la prueba de lo que le cuesta al profesor hacer la comparación de 14 trabajos. Como la matriz de comparación es simétrica, el número máximo de comparaciones si se comparan todos con todos es $N \times (N-1)/2$. En nuestro caso, 91 pares de

trabajos diferentes. Sin embargo, la aplicación aprovecha la información que se va generando y puede presentar una solución de ordenación sin necesidad de compararlos todos (la matriz de comparación completa tiene un alto grado de redundancia que el algoritmo aprovecha para estimar las comparaciones que puede calcular, en base a los comparaciones que se van incorporando). En la experiencia realizada, se usaron 3 criterios para comparar los trabajos. En la comparación número 30 (en torno a 20 minutos invertidos por el profesor) el sistema ya podía hacer una ordenación relativa de los 14 trabajos (una clasificación diferente para cada uno de los criterios). En la *Figura 3* mostramos la pantalla de gestión de la aplicación para el profesor, donde se pueden analizar los resultados con 30 comparaciones realizadas.

Panel de Control: MUGESP-31986-2015-16

Panel que permite la visualización de estadísticas de la actividad, link al comparador y acceso otras actividades de el usuario

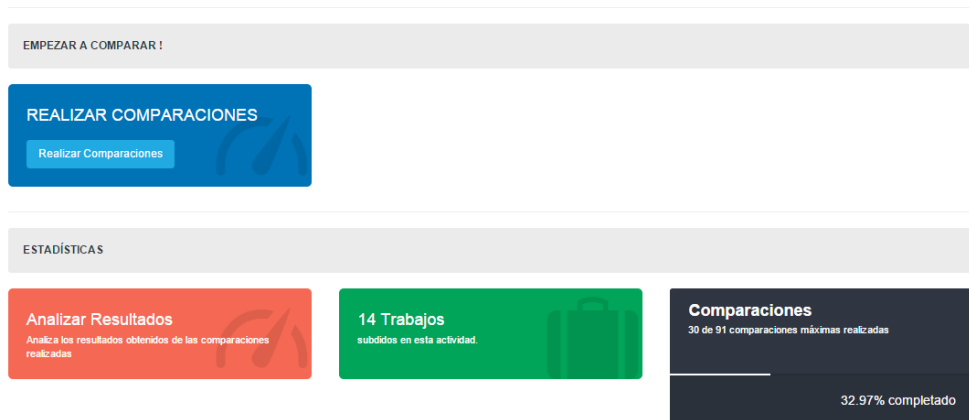


Figura 3. panel de control del profesor

Las *Figura 4*, *Figura 5*, *Figura 6* y *Figura 7* presentan la ordenación de los trabajos global (% de comparaciones “ganadas”) con datos de las comparaciones de los alumnos y la ordenación para cada uno de los tres criterios utilizando los datos de evaluaciones del profesor. En el criterio que valoraba la innovación de los dibujos, estos se clasificaron en 8 grupos (de mayor a menor innovación). Por ejemplo, un trabajo (el 14) destacaba sobre el resto, mientras que el grupo siguiente lo componían 3 trabajos (el 9, 8 y 5) de igual innovación pero menor que el trabajo 14 y mejores que el trabajo 12. En el criterio de pensamiento crítico solo había cuatro grupos de trabajo, el mejor con 4 dibujos, todos ellos equivalentes en cuanto a pensamiento crítico, pero mejores que cualquiera de los trabajos restantes. En la clasificación de iniciativa había seis grupos diferenciados de trabajos en cuanto a la intensidad de este criterio.

ORDEN TRABAJOS		OPCIONES ALGORITMO	
Nombre	Wissata	#Profundidad	#Anchura
ConstanzaTroianoFug.jpg	0.891	<input type="checkbox"/> Sweep	<input type="checkbox"/> Insert
image1.JPG	0.718	Replications: <input type="text" value="1"/>	<input type="checkbox"/> Reverse
JoaquinSanchezPlaneles.JPG	0.569	<input type="button" value="Ejecutar"/>	
AdrianVicentePanos.jpg	0.579		
IMG_20160218_144754.jpg	0.471		
AntonioFedericoRomeroGadea.jpg	0.304		
IanaKhuchenko.jpg	0.061		

Figura 4. Ordenación global de los trabajos por % de comparaciones ganadas

Resultados: MUGESP-31986-2015-16

Analiza y exporta los resultados obtenidos de las comparaciones que has realizado para esta actividad.

Innovacion

Análisis completo! Tenemos los trabajos ordenados para este criterio.

Grupo	Nombre
8	Orta 31982-mugesp2015 14
7	Orta 31982-mugesp2015 9 Orta 31982-mugesp2015 8 Orta 31982-mugesp2015 5
6	Orta 31982-mugesp2015 12
5	Orta 31982-mugesp2015 3
4	Orta 31982-mugesp2015 6 Orta 31982-mugesp2015 13
3	Orta 31982-mugesp2015 2
2	Orta 31982-mugesp2015 4 Orta 31982-mugesp2015 11 Orta 31982-mugesp2015 10 Orta 31982-mugesp2015 7
1	Orta 31982-mugesp2015 7



Figura 5. Ordenación para el criterio innovación

Critical thinking

Análisis completo! Tenemos los trabajos ordenados para este criterio.

Grupo	Nombre
4	Orta 31982-mugesp2015 3 Orta 31982-mugesp2015 13 Orta 31982-mugesp2015 8 Orta 31982-mugesp2015 12
3	Orta 31982-mugesp2015 2 Orta 31982-mugesp2015 11 Orta 31982-mugesp2015 4 Orta 31982-mugesp2015 5 Orta 31982-mugesp2015 8 Orta 31982-mugesp2015 9 Orta 31982-mugesp2015 10 Orta 31982-mugesp2015 7
2	Orta 31982-mugesp2015 14
1	Orta 31982-mugesp2015 7



Figura 6. Ordenación para el criterio pensamiento crítico

Iniciativa

Análisis completo! Tenemos los trabajos ordenados para este criterio.

Grupo	Nombre
6	Orta 31982-mugesp2015 13 Orta 31982-mugesp2015 14 Orta 31982-mugesp2015 3 Orta 31982-mugesp2015 5 Orta 31982-mugesp2015 6 Orta 31982-mugesp2015 12
5	Orta 31982-mugesp2015 2 Orta 31982-mugesp2015 11 Orta 31982-mugesp2015 10 Orta 31982-mugesp2015 7
4	Orta 31982-mugesp2015 9
3	Orta 31982-mugesp2015 8
2	Orta 31982-mugesp2015 4
1	Orta 31982-mugesp2015 7

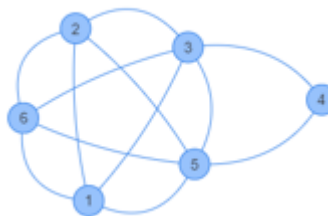


Figura 7. Ordenación para el criterio iniciativa

5. Conclusiones

Mediante la herramienta PASSAM es posible hacer de una manera cómoda y relativamente eficiente la evaluación de objetos por medio de comparación pareada. La experiencia que presentamos en este trabajo usa como objetos los dibujos entregados por los alumnos, pero se podría usar de manera similar para comparar trabajos escritos, soluciones a problemas o incluso las competencias transversales de los alumnos.

También hemos comprobado que el uso de la herramienta facilita la recolección y procesamiento de la información generada cuando se implica a los alumnos en la evaluación por pares, de modo que se pueden ver los resultados casi en tiempo real y utilizarlo como herramienta motivadora para que los alumnos realicen las tareas previas a las clases.

A pesar de las prometedoras posibilidades de PASSAM, este trabajo no está exento de limitaciones que queremos ir superando en investigación futura, replicando y ampliando la muestra de experiencias. Además, sería deseable comprobar en el futuro las siguientes cuestiones:

- a) La estabilidad de las soluciones si se siguen aportando datos de comparación una vez superado el punto donde es posible calcular una primera solución.
- b) Si el esfuerzo dedicado en evaluar, el acierto de las puntuaciones (medido contra un “gold-standard”) o la consistencia inter-rater e intra-rater, es superior cuando se emplea PASSAM (comparación pareada) en lugar de usar escalas tradicionales o rúbricas. Tanto en el caso de profesores noveles, como en el caso de alumnos o profesores experimentados.
- c) La consistencia de las ordenaciones realizadas con los datos proporcionados por los alumnos cuando se comparan con las de los profesores.
- d) Si es posible aplicarlo en grupos numerosos.
- e) Si puede emplearse para evaluar las competencias transversales.
- f) De qué depende (disparidad u homogeneidad de la calidad de los trabajos o la cantidad de trabajos) el número de comparaciones a realizar antes de alcanzar el punto en que se puede presentar la primera clasificación
- g) Métodos para trasladar la clasificación relativa a una nota de 0 a 10

6. Agradecimientos

Este trabajo ha sido realizado con la financiación de la Unión Europea [“FINCODA” proyecto 554493-EPP-1-2014-1-FI-EPPKA2-KA] (The European Commission support for the production of this publication does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein); y de la Universitat Politècnica de València PIME/2015/A/009/A “Evaluación de los indicadores del

comportamiento innovador en el alumno universitario” y (PIME/2015/A/014/A) “Peer ranking: La evaluación por comparación en las aulas”.

7. Referencias

- COMPONATION, P. J. and FARRINGTON, P. A. (2000) Identification of effective problem-solving tools to support continuous process improvement teams. *Engineering Management Journal*, 12(1), 23.
- DE LA HOZ I CASAS, J. and DE BLAS DEL HOYO, A. (2009) 'Learning by doing' methodology applied to the practical teaching of electrical machines. *International Journal of Electrical Engineering Education*, 46(2), 133-149.
- DOLAN, S. L., VALLE CABRERA, R., JACKSON, S. E. and SCHULER, R. S. (2007) *La gestión de los recursos humanos. Cómo atraer, retener y desarrollar con éxito el capital humano en tiempos de transformación*, (Madrid, Mcgraw-Hill).
- DOWNING, K., KWONG, T., CHAN, S. W., LAM, T. F. and DOWNING, W. K. (2009) Problem-based learning and the development of metacognition. *Higher Education*, 57(5), 609-621.
- GIBBINGS, P. and BRODIE, L. (2008) Assessment strategy for an engineering problem-solving course. *International Journal of Engineering Education*, 24(1), 153-161.
- HANRAHAN, S. J. and ISAACS, G. (2001) Assessing Self- and Peer- Assessment: The Students' Views. *Higher Education Research & Development*, 20(1), 53-70.
- HANSEN, P. K. (2003) Does productivity apply to PBL methods in engineering education? *International Journal of Engineering Education*, JAM-PDF(1), 177-182.
- LLORET, J. and MARIN-GARCIA, J. A. (2008) Comparing Novel and Stable Lecturers' Point of View when They Use University Students Working Groups in their Classrooms. *WSEAS Transactions on Advances in Engineering Education*, 11(5), 699-708.
- MARIN-GARCIA, J. A. (2007) Conversión de una asignatura hacia metodologías activas. Carga de trabajo de alumnos y profesores. *Revista Alternativas - Serie: Espacio Pedagógico*, 12(46/47), 51-60.
- MARIN-GARCIA, J. A. (2009) Los alumnos y los profesores como evaluadores. Aplicación a la calificación de presentaciones orales. *Revista Española De Pedagogía*, 67(242), 79-97.
- MARIN-GARCIA, J. A. and CONCHADO-PEIRÓ, A. (2012) Influencia de los Perfiles Docentes en el Uso y Preferencia de Dinámicas de Trabajo en Grupo con Alumnos Universitarios. *Formación Universitaria*, 5(5), 3-14.
- MARIN-GARCIA, J. A., GARCIA-SABATER, J. J., MAHEUT, J., VALERO-HERRERO, M. and ANDRES-ROMANO, C. (2012) *Gestión de recursos humanos para ingenieros de la rama industrial*, (Harlow, Pearson Education).
- MARIN-GARCIA, J. A., MIRALLES INSA, C., GARCIA-SABATER, J. J. and VIDAL CARREAS, P. I. (2008) Teaching management based on students teamwork: advantages, drawbacks and proposals for action. *Intangible Capital*, 4(2), 143-165.
- MARIN-GARCIA, J. A., RAMIREZ BAYARRI, L. and ANDREU ANDRÉS, M. A. (2015a) Comparación de los métodos de escalas y frecuencia de comportamiento para valorar la competencia de innovación. El punto de vista de alumnos y profesor en el caso de una asignatura de máster. *Congreso Nacional de Innovación Educativa y Docencia en Red- Universitat Politècnica de València-Valencia 30/06/15 al 1/07/15.*
- MARIN-GARCIA, J. A., RAMIREZ BAYARRI, L. and ATARES-HUERTA, L. (2015b) Protocol: Comparing advantages and disadvantages of Rating Scales, Behavior Observation Scales and Paired Comparison Scales for behavior assessment of competencies in workers. A systematic literature review. *Working Papers on Operations Management*, 2(6), 49-63.

- NORTHWOOD, M. D., NORTHWOOD, D. O. and NORTHWOOD, M. G. (2003) Problem-based learning (PBL): from the health sciences to engineering to value-added in the workplace. *Global Journal of Engineering Education*, JAM-PDF(2), 157-164.
- PRINS, F. J., SLUIJSMANS, D. M. A., KIRSCHNER, P. A. and STRIJBOS, J. W. (2005) Formative peer assessment in a CSCL environment: a case study. *Assessment & Evaluation in Higher Education*, 30(4), 417-444.
- RUBIN, B., FERNANDES, R., AVGERINOU, M. D. and MOORE, J. (2010) The effect of learning management systems on student and faculty outcomes. *Internet and Higher Education*, 13(1-2), 82-83.
- SPRONKEN-SMITH, R. and HARLAND, T. (2009) Learning to teach with problem-based learning. *Active Learning in Higher Education*, 10(2), 138-153.
- STEEL, C. (2009) Reconciling university teacher beliefs to create learning designs for LMS environments. *Australasian Journal of Educational Technology*, 25(3), 399-420.
- STEFANI, L. A. J. (1994) Peer, Self and Tutor Assessment - Relative Reliabilities. *Studies in Higher Education*, 19(1), 69-75.
- WOODS (1996) *Problem-based Learning: helping your students gain the most from PBL*.
- ZHANG, B., JOHNSTON, L. and KILIC, G. B. (2008) Assessing the reliability of self- and peer rating in student group work. *Assessment & Evaluation in Higher Education*, 33(3), 329-340.