

## **Análisis del grado de consistencia entre diferentes modos de evaluar las competencias transversales de creatividad, pensamiento crítico, liderazgo y trabajo en equipo/red<sup>1</sup>**

**Marin-Garcia, Juan A.<sup>a</sup>, Maheut, Julien<sup>b</sup>, Garcia-Sabater, Julio J.<sup>c</sup>**

<sup>a,b,c</sup> ROGLE. Dpto. de Organización de Empresas. Universitat Politècnica de València. Camino de Vera S/N 46021 Valencia. <sup>a</sup> [jamarin@omp.upv.es](mailto:jamarin@omp.upv.es), <sup>b</sup> [julien.maheut@gmail.com](mailto:julien.maheut@gmail.com) and <sup>c</sup> [jugarsa@omp.upv.es](mailto:jugarsa@omp.upv.es).

---

### **Abstract**

*In this paper, we analyze three different ways of qualifying soft skills in a Master course with 21 students. Both the holistic qualification of the 4 competences together and the analytical qualification in paired comparison present an excellent degree of agreement. However, the analytical qualification based on the qualification of exercises and tasks, despite reproduce quite well the order of the qualifications with respect to the judgment of the expert (teacher) has very different absolute values that indicate that the grades are concentrated and do not adequately represent the differences in the competences between the students*

**Keywords:** marks; grade; student performance; higher education; summative assessment; soft skills; paired comparison

---

### **Resumen**

*En este trabajo analizamos tres modos diferentes de calificar las competencias transversales en una asignatura de Máster con 21 alumnos. Tanto la calificación holística de las 4 competencias juntas, como la calificación analítica usando comparación pareada presentan un excelente grado de acuerdo. Sin embargo, la calificación analítica basada en calificación de ejercicios y tareas, pese a respetar el orden de las calificaciones respecto al juicio del experto (profesor) tiene unos valores absolutos muy dispares que indican que las notas se concentran y no representa adecuadamente las diferencias en cuanto a competencias entre los alumnos.*

**Palabras clave:** calificación; desempeño de estudiantes; educación universitaria; evaluación sumativa; competencias transversales; comparación pareada

---

<sup>1</sup> Esta comunicación puede considerarse un paso previo de otra publicación de los autores (Marin-Garcia et al., 2017a) cuyo objetivo, contribución y marco teórico son diferentes y amplían de manera significativa los objetivos de esta comunicación.

## **1. Introducción**

Las competencias transversales (*soft skills*) forman parte de la educación de estudiantes de ingeniería, no sólo por tratarse de requerimientos de agencias nacionales o internacionales (ANECA, EURACE, ABET), sino porque permiten equipar a los alumnos con herramientas que necesitarán en su trabajo como profesionales (Andersen, 2004; Anthony & Garner, 2016; Halfhill & Nielsen, 2007; Perello-Marin et al., 2016).

Sin embargo, la evaluación de estas competencias no es una tarea fácil (Halfhill & Nielsen, 2007; Ingols & Shapiro, 2014), porque se trata de conceptos multidimensionales (Balasundaram Ramakishnan & Ramadoss, 2009; Villa Sánchez & Poblete, 2007; Watts et al., 2006), no abundan los instrumentos validados (Marin-García & Santandreu-Mascarell, 2015) y existen varias opciones de escalas de respuesta para cada instrumento (Marin-García et al., 2015).

A pesar de la dificultad, es una tarea exigida en algunas universidades para tener los datos necesarios a la hora de confeccionar el expediente académico de los alumnos. En la Universitat Politècnica de València esta iniciativa se concreta en las asignaturas punto de control y queda reflejada en las guías docentes.

En este contexto, parece que las calificaciones de las competencias transversales no son tan fiables como se desearía. Al menos esa es la impresión, usando como fuente las reuniones de las Comisiones Académicas de varios Títulos de grado/master o conversaciones con profesores. La sensación de los profesores es que no están convencidos de las calificaciones otorgadas. Por otra parte, tomando los datos de cada asignatura, las calificaciones de las competencias transversales presentan poca dispersión entre los alumnos de un grupo y quizás inconsistencia entre diferentes asignaturas que ejercen de punto de control de la misma competencia.

Es posible que ambos fenómenos no sean un problema del sistema de calificación. Si todos los alumnos muestran unas competencias muy parecidas (porque las desarrollan hasta un nivel muy similar), es normal que las calificaciones estén concentradas en sólo unos pocos valores. Por otra parte, los alumnos pueden presentar comportamientos muy diferentes en asignaturas distintas, bien sea por la motivación de la temática, el método de enseñanza/aprendizaje, el profesorado, el horario de la asignatura o por los compañeros con los que interactúen. De modo que, por ejemplo, un alumno puede mostrarse muy creativo, en una asignatura y muy poco en otra que esté siendo, al mismo tiempo, punto de control de esta competencia. En esta situación, que el mismo alumno, el mismo año, reciba calificaciones diferentes en distintas asignaturas punto de control, no sería un problema de calificación; sino que el alumno se comporta de manera diferente en las asignaturas. Esto es algo que no nos extraña en las competencias específicas y, sin embargo, puede parecer raro en competencias transversales, quizás porque asumamos que son aspectos de personalidad que no son afectados por el contexto. Asunción que probablemente es errónea.

En la literatura existen algunas rúbricas o herramientas de diagnóstico que intentan mejorar la fiabilidad o validez de la calificación de las competencias transversales de

creatividad/innovación (Marin-Garcia et al., 2016a; Villa Sánchez & Poblete, 2007), pensamiento crítico (Marin-Garcia et al., 2016a; Villa Sánchez & Poblete, 2007); trabajo en equipo (Marin-Garcia et al., 2016a; Perello-Marín et al., 2016; Pratten et al., 2014; Villa Sánchez & Poblete, 2007) y comunicación (Marin-Garcia et al., 2017b; Marin-Garcia et al., 2008; Villa Sánchez & Poblete, 2007). Adicionalmente, en la Universitat Politècnica de València (UPV) se han propuesto unas rúbricas como orientación para la evaluación de esas competencias (<http://www.upv.es/contenidos/COMPTRAN/>)

## 2. Objetivos

- Comparar tres modos de evaluar las competencias transversales:
  - Evaluación holística de 0 a 10 basada en superación de objetivos de aprendizaje
  - Evaluación analítica mediante comparación pareada y incluyendo niveles de referencia inspirados en las rúbricas de competencias transversales UPV
  - Evaluación analítica del producto y proceso de aprendizaje de los alumnos utilizando rúbricas, listas de comprobación o baremos específicos creados por el profesor responsable de la asignatura
- Analizar la consistencia y la viabilidad de las tres alternativas

## 3. Desarrollo de la innovación

“Los datos utilizados provienen del curso 2016-17 en una asignatura de master con 21 alumnos. La asignatura se imparte durante 13 semanas lectivas en sesiones de 3,5 horas (una sesión semanal)” (Marin-Garcia et al., 2017a). La asignatura es punto de control de las competencias:

- CT4: innovación, creatividad y emprendimiento
- CT6: trabajo en equipo y liderazgo
- CT8: comunicación efectiva
- CT9: pensamiento crítico

Un aspecto particular de nuestra experiencia es que las competencias analizadas, no son “transversales” para la asignatura, sino que constituyen la esencia de la misma y representan el 100% de los objetivos de aprendizaje y, por lo tanto, de la nota final de los alumnos.

A continuación describimos los tres modos de obtener las calificaciones que hemos usado en esta investigación (Marin-Garcia et al., 2017a):

- “SE0.1: calificación de 0 a 10 holística, basada en los resultados de aprendizaje superados por la persona, utilizando las anotaciones tomadas en observaciones en clase, las preguntas directas a los alumnos, las preguntas que formulan los alumnos y los errores conceptuales que manifiestan. Cada semana el profesor dedicaba unas dos horas y media de clase a esta actividad mientras los alumnos realizaban actividades en su presencia”



- “SE0.2: utilizando un método de comparación pareada (Marin-Garcia et al., 2016b) derivada de trabajos previos (Marin-Garcia et al., 2014; Marin-Garcia et al., 2015). El proceso se inicia comparando parejas de alumnos indicando quién, a juicio del profesor, había mostrado mejor desempeño en creatividad, pensamiento crítico, intra-emprendimiento, trabajo en equipo, trabajo en red (figura 1). Se añadieron 3 alumnos ficticios para las comparaciones con niveles de referencia (figura 2), cada uno de ellos representaban el nivel excelente (equivalente a un 10), medio (equivalente a un 7) y bajo (equivalente a un 3) de las competencias analizadas (basado en las rúbricas orientadoras de la UPV). Para calificar las comparaciones se ofrecían 3 opciones (la persona A tiene mayor rendimiento que B, igual rendimiento, la persona B tiene mayor rendimiento que A)”. En la Figura 1 mostramos la pantalla de calificación cuando se comparaban dos alumnos; y en la Figura 2 cuando se comparaba un alumno de la asignatura con la descripción de comportamientos que representaban el nivel de referencia. “El sistema utiliza un algoritmo de grafos para decidir cuáles son las parejas a mostrar. En lugar de ser necesarias las 210 comparaciones posibles (para 21 alumnos), el procedimiento ha convergido a una solución tras 50 comparaciones (ver Tabla 1). El tiempo invertido por el profesor han sido 10 minutos en total. El resultado muestra 10 grupos (cada uno con uno o varios alumnos). Los alumnos que comparten grupo, son equivalentes en cuanto a desempeño. La distancia entre los grupos es proporcional al desempeño en las competencias analizadas. Una vez identificados los nodos (grupos de alumnos con similares resultados de aprendizaje) se ha procedido a volver a escalar las distancias en una nota de 0 a 10. Para ello, se ha identificado el nodo que contenía a cada nivel de referencia y se le ha imputado la nota (10, 7 o 3) correspondiente (ver Tabla 2). Al resto de nodos se les ha asignado una nota numérica proporcional a la distancia con los nodos ya calificados”
- “SE1: es una media ponderada. Consta de 30 registros por cada alumno. Estos registros están agrupados (se calcula el promedio) en 6 categorías cuya ponderación para el cálculo de la nota final de la asignatura aparece reflejado en la Tabla 3. Se han elaborado rúbricas detalladas para puntuar cada una de las actividades que se contemplan en los registros. En algunos casos son listas de comprobación Sí/no y en otros casos son rúbricas analíticas con ejemplos de comportamientos o resultados en 3-5 niveles de desempeño. Los alumnos han dispuesto de una copia de las rúbricas desde el principio del curso y han sido explicadas con detalle antes del inicio de cada una de las actividades. En algunos casos se han utilizado con ejercicios de ejemplos para que los alumnos tuvieran información de cómo se emplearían en la calificación de sus trabajos. El desempeño excepcional en alguno de los trabajos daba origen a notas por encima de 10/10”

Entre la calificación SE0.1 y SE0.2 transcurrieron 15 días en los que el profesor ya no trabajaba en la asignatura, por lo que consideramos que esas calificaciones son independientes y no están sesgadas por la secuencia en la que se han realizado.

Figura 1. Comparación cuando se mostraban dos alumnos

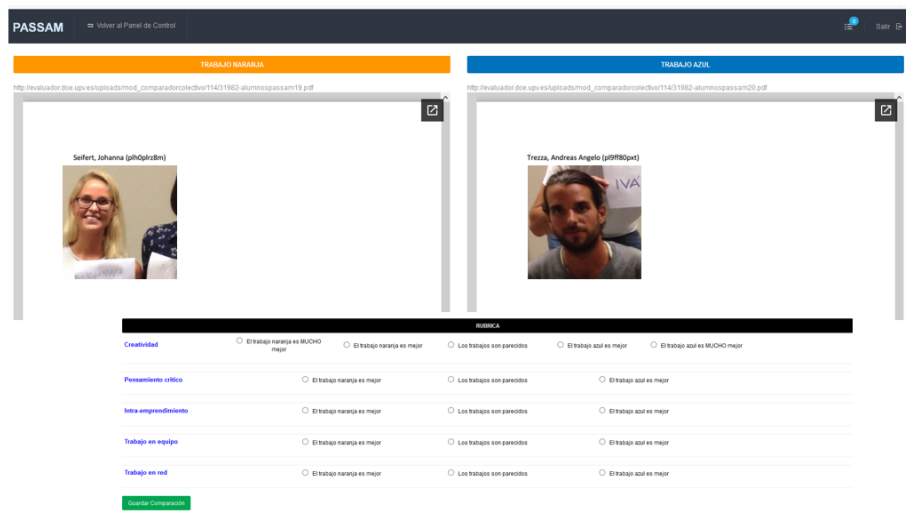
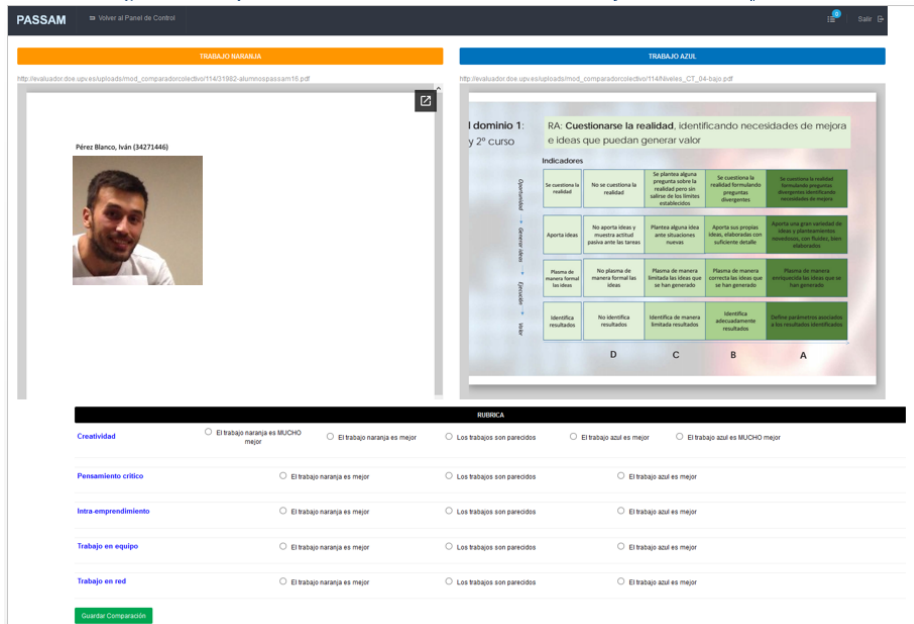


Figura 2. Comparación cuando se mostraba un alumno y un nivel de referencia



**4. Grupo Nombre**

10	<a href="#">A104</a>
9	A105 A113 A106
8	A107 A101
7	A119 A102
6	<a href="#">A121</a>
5	A114 A117
4	A111 A112
3	A109 A108 A118 A117
2	A110 A103 A115
1	<a href="#">A120</a>

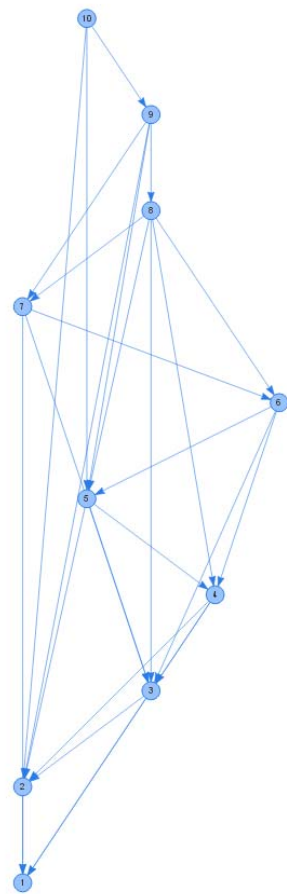


Tabla 1 . Nodos del paso 1 en SE0.2

Análisis completo! Tenemos los trabajos ordenados para este criterio.

Grupo	Nombre	Nota	Nota Interpolada
10		<input type="text" value="10"/>	10
9		<input type="text"/>	9.4
8		<input type="text"/>	8.8
7		<input type="text"/>	8.2
6		<input type="text"/>	7.6
5		<input type="text" value="7"/>	7
4		<input type="text"/>	6
3		<input type="text"/>	5
2		<input type="text"/>	4
1		<input type="text" value="3"/>	3

Tabla 2 . Re-escalado de notas. Paso 2 en SE0.2

Categoría	Nº de registros	Nota media de la clase	Peso (%)
<b>Asistencia (S)</b>	13	8.5	24%
<b>Diario y tareas en el Foro (V)</b>	2	6.9	20%
<b>Exámenes On-line (Y)</b>	3	9.4	17%
<b>Producto individual (T)</b>	3	9.0	18%
<b>Proyecto de mejora en equipo -Proceso de grupo (X)</b>	2	10.0	14%
<b>Puntos profesor (Z) (incluyendo rubrica de presentaciones orales)</b>	1	8.7	7%

Tabla 3 . Componentes del sistema de evaluación SE1

Nuestro objetivo es comparar SE1, SE0.1 y SE0.2 Consideraremos el ajuste como el coeficiente de correlación de rangos de Spearman y el coeficiente de correlación intra-clase (ICC(2,1) *agreement*) (Hair et al., 1995; Losilla et al., 2005; Marin-Garcia, 2009; Marin-Garcia et al., 2014; Marin-Garcia et al., 2017b).

## 5. Resultados

En la Tabla 4 resumimos los resultados de las calificaciones finales.

<b>PERSONA</b>	<b>SE0.1</b>	<b>SE0.2</b>	<b>SE1</b>
<b>AL01</b>	8	8.8	9.1
<b>AL02</b>	9.5	8.2	9.3
<b>AL03</b>	4	4	8.3
<b>AL04</b>	10	10	10.8
<b>AL05</b>	9.8	9.4	9.6
<b>AL06</b>	9.5	8.8	9.6
<b>AL07</b>	9.5	8.8	10.1
<b>AL08</b>	6.5	5	7.9
<b>AL09</b>	6	5	7.8
<b>AL10</b>	4	4	8.1
<b>AL11</b>	7	6	7.9
<b>AL12</b>	6	6	8.2
<b>AL13</b>	8.5	9.4	9.1
<b>AL14</b>	8	7	9.4
<b>AL15</b>	4	4	8.0
<b>AL16</b>	7	7	7.1
<b>AL17</b>	5	5	7.8
<b>AL18</b>	5	5	8.0
<b>AL19</b>	9	8.2	9.6
<b>AL20</b>	4	3	7.6
<b>AL21</b>	7	7.6	10.0

Tabla 4 . Calificaciones finales de las 21 personas matriculadas, dependiendo del sistema de cómputo elegido

Las correlaciones de rangos entre los diferentes métodos son todas elevadas (Tabla 5). Esto indica que, independientemente del valor absoluto de la calificación, todos los métodos ordenan de manera bastante similar a los estudiantes. Sin embargo, nosotros estamos interesados no en la correlación, sino en el acuerdo entre puntuaciones, ya que es la puntuación absoluta la que genera la categoría de calificaciones en el acta de la asignatura y queremos ver qué método genera unas calificaciones similares a las del evaluador experto.

Observando los valores de ICC(2,1) podemos comprobar que el método SE0.1 tienen un grado de acuerdo excelente con SE0.2. Es decir, el juicio del experto tiene una fiabilidad elevada tanto al cambiar el método para realizar la calificación, como al dejar pasar el tiempo y hacer un Test/re-test. Si tomamos las calificaciones SE0.2 como “gold standard”, se manifiesta un elevado desacuerdo con las calificaciones obtenidas con la media aritmética ponderada (SE1).



	SE0.1	SE0.2	SE1
SE0.1	-	<b>0.940</b>	
SE0.2	0.950**	-	<b>0.354</b>
SE1	0.727**	0.733**	-

Tabla 5 .Grado de acuerdo entre las medidas. En la diagonal inferior correlación Spearman's rho. En la submatriz superior (negrita) las ICC(2,1) de cada variable comparada con SE0.2. \*\* nivel de significación 1%

## 6. Conclusiones

Presentamos los resultados de comparar varios modos de calificar competencias transversales. La principal contribución para los profesores se aplica a grupos poco numerosos donde se puede llegar a conocer con detalle el grado de aprendizaje de los alumnos, el método de comparación pareada es el que presenta unas calificaciones más acordes con el juicio experto del profesor. La comparación pareada se hace de una manera muy eficiente en tiempo. Requiriendo solo 10 minutos adicionales respecto a las horas dedicadas a observación y *feedback* formativo realizado durante el curso. La comparación pareada es un método muy poco usado en la evaluación de estudiantes y convendría seguir investigando sobre las ventajas e inconvenientes de este método, que parece prometedor, con los datos del caso estudiado en este artículo.

La principal limitación de este trabajo es que sólo se ha probado en un grupo. Sería recomendable replicar la investigación con asignaturas similares y asignaturas con grupos masificados, de otros niveles (grado con 70-90 alumnos por grupo) y de otras titulaciones.

## Agradecimientos

Este trabajo ha sido parcialmente financiado por la Universitat Politècnica de Valencia (PIME/2016/A/027/A) "La evaluación pareada como metodología para la evaluación del pensamiento crítico de los alumnos".

## 7. Referencias

- Andersen, A. (2004). Preparing engineering students to work in a global environment to cooperate, to communicate and to compete. *European Journal of Engineering Education*, 29(4), 549-558.
- Anthony, S., & Garner, B. (2016). Teaching soft skills to business students: An analysis of multiple pedagogical methods. *Business and Professional Communication Quarterly*, 79(3), 360-370. doi:10.1177/2329490616642247
- Balasundaram Ramakishnan, S., & Ramadoss, B. (2009). Assessment using multi-criteria decision approach for higher order skills learning domains. *International Journal on E-Learning*, 8(2), 241-262.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis* (4<sup>o</sup> ed.). New Jersey: Prentice Hall.



- Halfhill, T. R., & Nielsen, T. M. (2007). Quantifying the "softer side" of management education: An example using teamwork competencies. *Journal of Management Education, 31*(1), 64-80.
- Ingols, C., & Shapiro, M. (2014). Concrete steps for assessing the soft skills in an mba program. *Journal of Management Education, 38*(3), 412-435.
- Losilla, J. M., Navarro, J. B., Palmer, A., Rodrigo, M. F., & Ato, M. (2005). *Análisis de datos. Del contraste de hipótesis al modelado estadístico*. Barcelona: Edicions a Petició.
- Marin-Garcia, J. A. (2009). Los alumnos y los profesores como evaluadores. Aplicación a la calificación de presentaciones orales. *Revista Espanola De Pedagogia, 67*(242), 79-97.
- Marin-Garcia, J. A., Andreu Andres, M. A., Atares-Huerta, L., Aznar-Mas, L. E., Garcia-Carbonell, A., González-Ladrón-de-Guevara, F., . . . Watts, F. (2016a). Propossal of framework for innovation competencies development and assessment (fincoda). *WPOM-Working Papers on Operations Management, 7*(2), 119-126. doi:<http://dx.doi.org/10.4995/wpom.v7i2.6472>
- Marin-Garcia, J. A., Aragonés Beltran, P., & Melón, G. (2014). Intra-rater and inter-rater consistency of pair wise comparison in evaluating the innovation competency for university students. *WPOM-Working Papers on Operations Management, 5*(2), 24-46. doi:<http://dx.doi.org/10.4995/wpom.v5i2.3220>
- Marin-Garcia, J. A., Garcia-Sabater, J. P., Morant Llorca, J., & Conejero, J. A. (2016b). *Passam: Peer assessment and monitoring system*. Paper presented at the Congreso Nacional de Innovación Educativa y Docencia en Red- Universitat Politècnica de València-Valencia 07/07/16 al 08/07/16.
- Marin-Garcia, J. A., Maheut, J., & Garcia-Sabater, J. J. (2017a). Comparison of different ways of compute the grades of continuous assessment in the final grade. *WPOM-Working Papers on Operations Management, 8*(SpIss), in press.
- Marin-Garcia, J. A., Maheut, J., & Santandreu-Mascarell, C. (2017b). Protocol: Inter-rater and intra-rater consistency validation of a rubric to assess oral presentation skills for university students. *WPOM-Working Papers on Operations Management, 7*(2), (in press).
- Marin-Garcia, J. A., Miralles Insa, C., & Marin Garcia, P. (2008). Oral presentation and assessment skills in engineering education. *International Journal of Engineering Education, 24*(5), 926-935.
- Marin-Garcia, J. A., Ramirez Bayarri, L., & Atares-Huerta, L. (2015). 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. *WPOM-Working Papers on Operations Management, 2*(6), 49-63. doi:<http://dx.doi.org/10.4995/wpom.v6i2.4032>
- Marin-Garcia, J. A., & Santandreu-Mascarell, C. (2015). What do we know about rubrics used in higher education? *Intangible Capital, 11*(1), 118-145. doi:<http://dx.doi.org/10.3926/ic>.
- Perello-Marin, M. R., Vidal-Carreras, P. I., & Marin-Garcia, J. A. (2016). What do undergraduates perceive about teamwork? *International Journal of Engineering Education, 32*(3), 1171-1181.
- Pratten, M. K., Merrick, D., & Burr, S. A. (2014). Group in- course assessment promotes cooperative learning and increases performance. *Anatomical Sciences Education, 7*(3), 224-233. doi:10.1002/ase.1397
- Villa Sánchez, A., & Poblete, M. (2007). *Aprendizaje basado en competencias. Una propuesta para la evaluación de las competencias genéricas*. Bilbao: Universidad de Deusto.

Watts, F., García-Carbonell, A., & Llorens, J. (2006). Introducción a la evaluación compartida: Investigación multidisciplinar. In F. Watts & A. García-Carbonell (Eds.), *La evaluación compartida: Investigación multidisciplinar* (1 ed., pp. 1-9). Valencia: Editorial de la UPV

