



## 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

Juan A. Marin-Garcia<sup>a</sup>, Lucia Ramirez Bayarri<sup>b</sup> y M<sup>a</sup> Angeles Andreu Andres<sup>c</sup>

<sup>a</sup> ROGLE. Dpto. de Organización de Empresas. Universitat Politècnica de València. Camino de Vera S/N 46021 Valencia. [jamarin@omp.upv.es](mailto:jamarin@omp.upv.es), <sup>b</sup>Department d'Ensenyament de la Generalitat de Catalunya. [luraba@gmail.com](mailto:luraba@gmail.com), <sup>c</sup>Departamento de Idiomas. Universitat Politècnica de València. [maandreu@idm.upv.es](mailto:maandreu@idm.upv.es)

---

### Abstract

*Skills assessment is a complex task that includes deciding on various aspects or dimensions of competencies. Moreover, it is not always easy for teachers to observe and assess certain skills. It is therefore recommended sometimes incorporate students as evaluators of their own abilities. Self-assessment has benefits because it increases the critical observation and student capacity. Similarly, it encourages reflection and meta-cognition, which are two important components in independent learning throughout life. However, despite the benefits cited, students supposed them a challenge, assigning a rating using a score on a scale of 0-10, not all students are willing to participate as evaluators, nor are they equally reliable. This research compares two different scales for the collection of data for self-assessment of competence of innovation and teamwork with 16 master students. We base or comparison on the degree of agreement between the scores on both scales. We also use the qualitative opinions of the students.*

**Keywords:** *self-assessment of students; higher education; self-evaluation; innovation competency*

---

### Resumen

*La evaluación de competencias es una tarea compleja que incluye decidir sobre varios aspectos o dimensiones de la competencia. Por otra parte, no siempre es fácil que los profesores puedan o sepan observar y valorar determinadas habilidades. Por ello, se ha recomendado, en ocasiones, incorporar a los alumnos como evaluadores de sus propias habilidades. La auto-evaluación presenta beneficios pues aumenta la capacidad crítica y de observación de los estudiantes. Del mismo modo, fomenta la reflexión y la meta-cognición, que son dos componentes importantes en el aprendizaje autónomo a lo largo de la vida. No obstante, a pesar de los beneficios citados,*

*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*

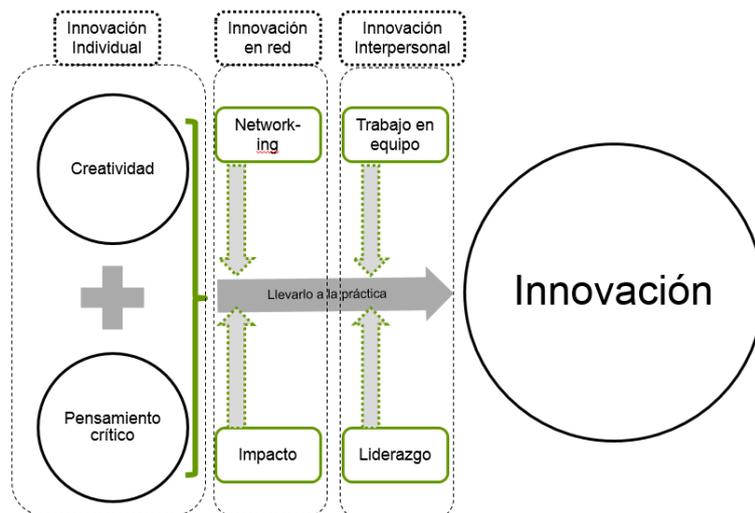
*a los estudiantes les supone un reto, el asignar una nota usando una puntuación en una escala de 0 a 10 y no todos los alumnos están dispuestos a participar como evaluadores, ni son igual de fiables cuando evalúan. Esta investigación compara dos escalas diferentes para la recogida de datos para la auto-evaluación de la competencia de innovación y trabajo en equipo de los 16 estudiantes de una asignatura de master basándose en el grado de acuerdo de las puntuaciones obtenidas por ambas escalas y la valoración crítica de los estudiantes.*

**Palabras clave:** *Evaluación de estudiantes; educación universitaria; auto-evaluación; competencia de innovación; comparación de métodos*

---

## Introducción

Una de las competencias más demandadas en las sociedades avanzadas es la competencia de innovación, que aparece reflejada en la mayoría de los planes de estudios de España bajo diferentes nombres: innovación, creatividad, emprendimiento, habilidad para tomar decisiones, capacidad para resolver problemas... (Andreu Andrés and García-Casas, 2014, Ingols and Shapiro, 2014, Marin-Garcia et al., 2008). En este trabajo nos hemos basado en el modelo de competencia de innovación representado en la Figura 1, que es una evolución de trabajos previos (Marin-Garcia et al., 2013, Marin-Garcia et al., 2010, Marin-Garcia et al., 2011).



**Figura 1.- Modelo de innovación (Fuente: Marin-Garcia (2015) sin publicar)**

La evaluación de competencias es una tarea compleja que incluye decidir sobre varios aspectos o dimensiones de la competencia. En general, para la mayoría de las competencias profesionales, no suele haber demasiado acuerdo acerca de qué subdimensiones (capacidades o habilidades) las componen y, por lo tanto, qué habilidades deben evaluarse (Lohmann and Prumper, 2006, Marin-Garcia et al., 2013).

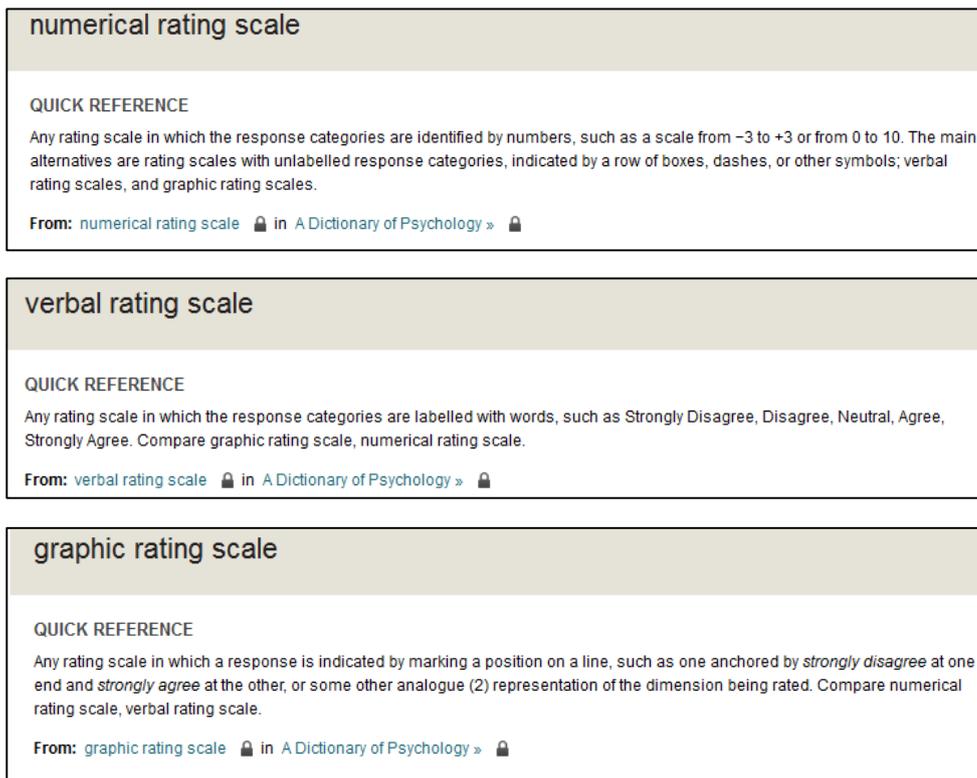
Por otra parte, no siempre es fácil que los profesores puedan o sepan observar y valorar determinadas habilidades a través de los comportamientos de sus estudiantes. La evaluación del comportamiento se centra en “identificar aspectos claramente observables y el modo en que la persona interactúa con su entorno” (Groth-Marnat, 2003). Los principales instrumentos para la evaluación del comportamiento son los cuestionarios y estrategias de observación (Dowdy et al., 2013). En esta comunicación, nos vamos a centrar en los primeros, en su modalidad de cuestionario autoadministrado por el propio estudiante que se autoevalúa (Groth-Marnat, 2003). Tanto la auto-evaluación como la evaluación de compañeros presenta beneficios pues aumenta la capacidad crítica y de observación de los estudiantes (Van Overveld and Verhoeff, 2013). Del mismo modo, fomenta la reflexión y la meta-cognición, que son dos componentes importantes en el aprendizaje autónomo a lo largo de la vida (Lind et al., 2002, Ljungman and Silén, 2008). No obstante, a pesar de los beneficios citados, a los estudiantes les supone un reto, el asignar una nota y no todos los alumnos están dispuestos a participar como evaluadores, ni son igual de fiables cuando evalúan (Pond, 2007, Van Overveld and Verhoeff, 2013).

Dentro de las diferentes tradiciones en las que se ha analizado la evaluación por observación (psicología clínica, pedagogía, gestión de recursos humanos), parece haber un consenso en que hay otro aspecto importante en el proceso de evaluación, además de quién hace la medición. Es el modo en que se mide (las escalas o instrumentos utilizados). No es nuestro objetivo en este trabajo hacer un resumen de las ventajas y limitaciones de las observaciones de comportamiento frente a las escalas de calificación, la frecuencia con las que se usan o las recomendaciones para la puesta en práctica en general de esta estrategia de evaluación. Mucha de esa información se pueden encontrar en Dowdy et al (2013) o en Marin-Garcia et al. (2012). Nosotros nos centraremos en describir los diferentes tipos de escalas que podemos usar para evaluar competencias por medio de cuestionarios autoadministrados y compararemos los resultados de dos de ellas.

Podemos considerar cuatro posibles métodos de recogida de los datos de entrada necesarios para la evaluación de los alumnos. Las listas ordenadas (rankings), las escalas de puntuación (ratings), las escalas basadas en observación de frecuencia de comportamientos (BARS y BOS) y la comparación pareada (*paired comparisons*) (Dowdy et al., 2013, Hatzinger and Dittrich, 2012, Marin-Garcia et al., 2012). En este trabajo nos vamos a centrar en el segundo y tercero de ellos.

*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*

Las escalas de puntuación se caracterizan por que las categorías de respuesta, compartidas por diferentes preguntas o ítems en el cuestionario, se identifican con números o distancia respecto a límites o anclas verbales (Figura 2).



**Figura 2: Escalas de evaluación**

(fuente: <http://www.oxfordreference.com/view/10.1093/oi/authority.20110803095904227?rskey=VE2Fjz&result=3>)

Por otra parte, las escalas de evaluación ancladas en comportamientos (BARS), identifican los aspectos claves de un puesto de trabajo o posición (por ejemplo, estudiante) y las posibles conductas que pueden mostrar al desarrollar sus tareas, ordenándolas desde las más ineficientes o indeseables hasta las más eficientes o deseables (Figura 3)

A technique for evaluating the performance of an employee which can be used as part of the appraisal process. The technique involves: (1) breaking down a particular job into its key tasks (performance dimensions); (2) identifying a range of possible behaviours that can be displayed by an employee in undertaking each task; (3) placing these behaviours on a scale ranging from ineffective performance to excellent performance; and (4) assessing the jobholder (employee) against these scales for each of the tasks. As a consequence, a profile of job performance is created for each employee covering the various aspects of his or her work. Typically, the process of devising the performance dimensions involves jobholders themselves, and this gives greater authenticity to the rating scales. [See behavioural observation scales (BOS).]

(1) breaking down a particular job into its key tasks (performance dimensions); (2) identifying a range of possible behaviours that can be displayed by an employee in undertaking each task; (3) placing these behaviours on a scale ranging from ineffective performance to excellent performance; and (4) assessing the jobholder (employee) against these scales for each of the tasks. [Show Less](#)

From: behaviourally anchored rating scales in A Dictionary of Human Resource Management »

**Figura 3.- Behaviourally anchored rating scales (BARS) (Fuente:**  
**<http://www.oxfordreference.com/view/10.1093/oi/authority.20110803095456792?rskey=cWbWx8&result=3>**)

Por último, las escalas de observación de comportamiento (BOS) identifican una serie de conductas observables en los alumnos, como en BARS, pero se recogen datos de la frecuencia en que los estudiantes muestran esas conductas para conseguir un adecuado rendimiento en las tareas encargadas durante el curso. Son especialmente adecuadas cuando se quiere evaluar el proceso no solo el resultado o la productividad de los estudiantes (Figura 4)

A technique for evaluating the performance of an employee which can be used as part of the appraisal process. Like behaviourally anchored rating scales, the BOS technique involves a process of identifying the key tasks for a particular job, but the difference is that employees are evaluated according to how frequently they exhibit the required behaviour for effective performance. The scores for each of these observed behaviours can then be totalled to produce an overall performance score. In such instances, the various measures of behaviour are normally weighted to reflect the relative importance of the measure to the overall job.

From: behavioural observation scales in A Dictionary of Human Resource Management »

**Figura 4.- Behavioural observation scale (BOS) (Fuente:**  
**<http://www.oxfordreference.com/view/10.1093/oi/authority.20110803095456852>**)

La ponderación, en las diferentes dimensiones, de los valores obtenidos en las escalas de puntuación o de las frecuencias de las alternativas de comportamientos observados permite obtener una calificación global, que puede usarse para ordenar a los estudiantes de mejor a peor rendimiento.

## Objetivos

El objetivo de esta investigación es comparar dos escalas diferentes para la recogida de datos para la auto-evaluación de la competencia de innovación y trabajo en equipo de los estudiantes de una asignatura de master basándose, por un lado, en el grado de acuerdo de las puntuaciones obtenidas por ambas escalas y por otro la valoración crítica de los estudiantes acerca de cuál de las dos opciones les ha resultado más sencillo de contestar si creen que los

*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*

resultados reflejan bien cómo son respecto a la innovación y al trabajo en equipo y si les ha ha resultado útil rellenar el cuestionario

## Desarrollo de la innovación

La asignatura donde se ha realizado la intervención es de un máster internacional oficial, su principal objetivo de aprendizaje está relacionado con las competencias de innovación y de trabajo en equipo, tiene 16 alumnos matriculados. La asignatura se impartió en inglés, siendo la mitad de los alumnos de habla castellana y la otra mitad proveniente de intercambio de diferentes países europeos (con especial incidencia de Alemania). Los datos fueron recogidos a principios de septiembre de 2014, por medio de una plataforma de encuestas web (limesurvey) siendo las preguntas mostradas en castellano e inglés. La propia plataforma web grababa también el tiempo invertido por los estudiantes para completar su auto evaluación.

La segunda semana de clase se solicitó a los alumnos que se autoevaluaran usando el cuestionario con 25 preguntas con formato escala de evaluación de 11 niveles de respuesta (de 0 a 10), pero con 6 anclas verbales (desde “no observado” as “excelente”). Se decidió usar anclas verbales y 11 niveles de respuesta basándonos en trabajos previos que avalan que las propiedades psicométricas de las escalas mejoran cuando se utilizan anclas verbales y cuando hay más de 7-9 niveles en la comparación (De Beuckelaer et al., 2013, Yiu et al., 2007), ya que proporcionar suficiente rango de variabilidad, sin complicar en exceso el proceso de comparación (ver Figura 5). Los 25 ítems se promedian para calcular las puntuaciones en 3 dimensiones de la innovación (individual -12 ítems-, interpersonal -8 ítems- y en red -5 ítems-).

	Not observed/ not demonstrated	Poor	Needs to improve	Pass	Good	Excellent
Points	0	1-2	3-4	5-6	7-8	9-10
Propongo nuevas maneras de poner en marcha las ideas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Anticipo cómo se pueden desarrollar los acontecimientos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Colaboro activamente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Contribuyo al buen funcionamiento del grupo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Añado impacto social a la tarea	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hago que los demás contribuyan en las tareas del grupo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Escucho a mis compañeros de grupo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Afronto los conflictos con flexibilidad con el fin de alcanzar acuerdos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Identifico las relaciones entre las diferentes partes o aspectos de la tarea	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Figura 5.- Extracto de las escala de valoración usada en la primera toma de datos (Watts et al., 2013, Watts et al., 2012)**

Una semana más tarde (tiempo que se consideró suficiente para evitar efecto recuerdo, pero al mismo tiempo no excesivo para garantizar que los alumnos no habían tenido tiempo para modificar su competencia de innovación), se les solicitó a los mismos alumnos que se volvieran a evaluar, pero usando una escala BOS. Este autodiagnóstico se componía de 18 grupos de preguntas derivadas de las 25 originales, cada una de ellas con 5 comportamientos

ordenados desde menos favorecedor de la conducta innovadora a más favorecedor. Las respuestas eran la frecuencia en la que se mostraba cada uno de los comportamientos. Las puntuaciones de los 5 comportamientos de cada pregunta se ponderaban para calcular la puntuación de la pregunta y, posteriormente, las 18 preguntas se agrupaban para calcular las puntuaciones de las 3 dimensiones de innovación (del mismo modo que se hace en la versión “escala de puntuación”). Se ha pasado a 18 grupos de preguntas, pues en las 25 preguntas originales había ítems que representaban comportamientos de diferentes niveles de una misma pregunta en BOS. En la versión BOS, están presentes las 25 preguntas originales como anclas de un determinado nivel de comportamiento favorecedor de la innovación. Es decir, ambos cuestionarios pretenden medir lo mismo, pero de diferente manera. En la Figura 6 se presentan los grupos de comportamientos asociados a dos de las preguntas “creatividad” e “impacto”.

Valora, en cada fila, con qué frecuencia te comportas de las siguientes maneras cuando formas parte de grupos que buscan soluciones u oportunidades (toma como referente grupos de trabajo. Si no estás trabajando, puedes usar los grupos para tareas como estudiante u otros grupos en los que participes –por ejemplo asociaciones, etc.-)

	0 Nunca me comporto así	1 Muy pocas veces me comporto así	2 Pocas veces me comporto así	3 Bastantes veces me comporto de esta manera	4 Muchas veces me comporto así	5 Casi siempre me comporto así	Sin respuesta
No apporto ideas	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Aporto ideas que tienen poco que ver con la tarea	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Propongo ideas adecuadas a las exigencias de la tarea	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Ofrezco ideas que son originales en contenido	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ofrezco nuevos modos de materializar las ideas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

Valora, en cada fila, con qué frecuencia te comportas de las siguientes maneras cuando formas parte de grupos que buscan soluciones u oportunidades (toma como referente grupos de trabajo. Si no estás trabajando, puedes usar los grupos para tareas como estudiante u otros grupos en los que participes –por ejemplo asociaciones, etc.-)

Por favor, seleccione la respuesta apropiada para cada concepto:

	0 Nunca me comporto así	1 Muy pocas veces me comporto así	2 Pocas veces me comporto así	3 Bastantes veces me comporto de esta manera	4 Muchas veces me comporto así	5 Casi siempre me comporto así
No me preocupa el impacto de las tareas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Intento que las tareas tengan algo de impacto para mi grupo o para mi organización	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Intento que las tareas tengan mucho impacto para mi grupo o para mi organización	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Intento que las tareas tengan algo de impacto en la sociedad	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Intento que las tareas tengan mucho impacto en la sociedad	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figura 6.- Extracto de la escala usada en la segunda toma de datos

*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*

Utiliza esta escala para valorar las preguntas que te hacemos a continuación. Responde indicando qué has demostrado en el contexto de tareas que te haya indicado el profesor como referente de este cuestionario (puede ser una actividad concreta de la asignatura -práctica o trabajo- o tu trabajo en la asignatura completa):

	No demostrado/ no observado	Necesito mejorarla mucho	Necesito mejorar	Suficiente/ aceptable	Buena	Excelente
Puntos	0	1-2	3-4	5-6	7-8	9-10

\*  
Por favor, seleccione la respuesta apropiada para cada concepto:

	1	2	3	4	5	6	7	8	9	10
¿Te consideras una persona innovadora? Valorar tu competencia para fomentar la innovación, considerando la innovación como un proceso que permite la creación y puesta en práctica de algo (producto o servicio, métodos, mercados, suministros, modelos de negocios, modelos de gestión o tareas en general) que sea percibido como novedad por la organización/grupo y que persiga una mejora del rendimiento o resultados de la organización/grupo	<input type="radio"/>									
Valora en qué medida tus comportamientos o habilidades te permiten, por ti misma, ser una persona innovadora en la ejecución de tareas	<input type="radio"/>									
Valora tu capacidad para hacer actuar a otros y potenciar la capacidad individual de innovar por medio de la interacción con el grupo responsable de completar un proyecto o tarea	<input type="radio"/>									
Valora tu capacidad para crear/emplear una red de contactos de modo que permitan al grupo (responsable de completar un proyecto o tarea) buscar soluciones adecuadas en un entorno más amplio del habitual	<input type="radio"/>									

**Figura 7.- Escalas monoitem generales de la innovación**

El grado de acuerdo de las puntuaciones de los alumnos entre las dos versiones se computará con los coeficientes de correlación y la prueba de muestras T pareadas (Hair et al., 1999).

El análisis cualitativo de las percepciones de los estudiantes está basado en las respuestas a estas preguntas abiertas:

- Cuál de las dos opciones, “escala de puntuación” (normal) o BOS, te ha resultado más sencillo de contestar y por qué.
- Una vez completado cada cuestionario te llega un informe con tu puntuación:
  - ¿Eran parecidos los resultados en las dos versiones?
  - ¿Crees que los resultados reflejan bien cómo eres respecto a la innovación y al trabajo en equipo? ¿En las dos opciones, sólo en una (cuál), ninguna? ¿Por qué?
- ¿Te ha resultado útil rellenar el cuestionario? ¿Cuál más? ¿Por qué?

El análisis se ha realizado siguiendo un protocolo de codificación sistemática soportado por el programa informático ATLAS-TI (Bringer et al., 2006, Friese, 2012).

## Resultados

El tiempo invertido en completar el autodiagnóstico está en torno a 5 minutos en la versión escalas de puntuación y de 18 minutos en la versión BOS (Tabla 1). Esto es debido a que el número de valoraciones que se hacen en la versión BOS es más elevado (18 preguntas, cada una de ellas con 5 comportamientos a valorar su frecuencia).

**Tabla 1.- Tiempo invertido en cada una de las versiones (minutos)**



	<i>N</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Mean</i>	<i>Std. Deviation</i>
InterviewTime	16	2,08	9,84	4,39	2,23
interviewtime_BOS	16	9,28	29,78	18,38	6,00

A continuación hemos analizado las medias y correlaciones entre las puntuaciones obtenidas en las dos tomas de dato. En principio, nuestra hipótesis era que ambas medidas debían ser iguales pues en una semana el alumno no ha cambiado sustancialmente los comportamientos asociados a la competencia evaluada. Los cuatro primeros pares de comparación se hacen con exactamente las cuatro mismas preguntas a modo de test-retest (Figura 7). Al alumno se le pide que se evalúe con una escala de 0 a 10 (no demostrado a excelente) su competencia de innovación en general (SingleItemInnov y SingleItemInnov\_BOS), su capacidad para la innovación individual (SingleItemIndiv y SingleItemIndiv\_BOS), su capacidad para la innovación en grupo (SingleItemInter y SingleItemInter\_BOS) , y su capacidad para la innovación a través de una red de contactos (SingleItemNetw y SingleItemNetw\_BOS). Los resultados (Tabla 2, Tabla 3 y Tabla 4) muestran que no hay diferencias significativas en las medias entre los dos momentos y que las correlaciones entre las medidas son altas (sobre todo en la capacidad de innovación interpersonal y en res). Por ello nos servirá como evidencia de que el alumno se percibe sin apenas cambios en cuanto innovación.

Veamos qué ocurre con las puntuaciones en las dos versiones de los cuestionarios (Pair 5 a Pair 7 en las tablas). En la dimensión individual, no hay diferencias significativas entre las puntuaciones de los dos tipos de cuestionarios, aunque la correlación es moderada. En las otras dos dimensiones, a pesar de que existen diferencias significativas, la correlación entre las medidas es elevada. En la versión BOS los alumnos se puntúan con menos puntos, pero la posición de las personas, de mayor a menor puntuación, se mantiene. Es decir, parece que la versión escala de puntuación, se satura más fácilmente que la versión BOS.

**Tabla 2.-Descriptivos**

		<i>Mean</i>	<i>Std. Deviation</i>
Pair 1	SingleItemInnov	6,81	1,471
	SingleItemInnov_BOS	6,19	1,834
Pair 2	SingleItemIndiv	7,13	1,360
	SingleItemIndiv_BOS	6,69	1,778
Pair 3	SingleItemInter	7,06	1,914
	SingleItemInter_BOS	7,13	1,784
Pair 4	SingleItemNetw	7,00	2,066
	SingleItemNetw_BOS	6,81	1,905
Pair 5	ICBIndividual	72,08	9,01
	ICBIndividual_BOS100	73,51	10,75
Pair 6	ICBInterpersonal	78,43	8,84
	ICBInterpersonal_BOS100	71,31	11,68
Pair 7	ICBRed	73,87	17,89
	ICBRed_BOS100	59,22	14,31

*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*

**Tabla 3.-correlaciones Pearson**

		<i>N</i>	<i>Correlation</i>	<i>Sig.</i>
Pair 1	SingleItemInnov & SingleItemInnov_BOS	16	,657	,006
Pair 2	SingleItemIndiv & SingleItemIndiv_BOS	16	,568	,022
Pair 3	SingleItemInter & SingleItemInter_BOS	16	,837	,000
Pair 4	SingleItemNetw & SingleItemNetw_BOS	16	,881	,000
Pair 5	ICBIndividual & ICBIndividual_BOS100	16	,546	,029
Pair 6	ICBInterpersonal & ICBInterpersonal_BOS100	16	,806	,000
Pair 7	ICBRed & ICBRed_BOS100	16	,762	,001

**Tabla 4.- Paired Samples Test (95% intervalo de confianza) 15 df; Sig 2 tailed**

		<i>Mean difference</i>	<i>Std. Deviation</i>	<i>Lower</i>	<i>Upper</i>	<i>t</i>	<i>Sig.</i>
Pair 1	SingleItemInnov - SingleItemInnov_BOS	,625	1,408	-,125	1,375	1,775	,096
Pair 2	SingleItemIndiv - SingleItemIndiv_BOS	,438	1,504	-,364	1,239	1,163	,263
Pair 3	SingleItemInter - SingleItemInter_BOS	-,063	1,063	-,629	,504	-,235	,817
Pair 4	SingleItemNetw - SingleItemNetw_BOS	,188	,981	-,335	,710	,764	,456
Pair 5	ICBIndividual - ICBIndividual_BOS100	-1,432	9,538	6,515	3,650	-,601	,557
Pair 6	ICBInterpersonal - ICBInterpersonal_BOS100	7,123	6,943	3,423	10,823	4,104	,001
Pair 7	ICBRed - ICBRed_BOS100	14,654	11,618	8,463	20,845	5,045	,000

En las contestaciones abiertas, se manifiesta un acuerdo de los estudiantes en que ambas versiones ofrecen unos resultados similares (Figura 8). Esos resultados son percibidos como muy válidos, porque la escala de 0 a 100 recibida coincide con su percepción del despliegue de las diferentes dimensiones de su competencia de innovación (Figura 9). También hay unanimidad en considerar el autodiagnóstico como una actividad muy útil en la asignatura, porque les ayuda a la reflexión de sus comportamientos; motiva a aprender y detectar áreas fuertes o áreas de mejora (Figura 10). Aparece en algunos casos una interesante solicitud de que no se limite a auto-diagnóstico sino que les parecería muy útil contar con la visión de terceras personas (Figura 11). Por último, todos coinciden en considerar que la versión “escala” es mucho más breve que la versión BOS pero eso no siempre significa que sea más fácil de contestarla. Para algunos alumnos, la versión BOS es más fácil de contestar, aunque sea más larga, porque es más descriptiva y les permite ser más precisos y claros en sus contestaciones. Además, algunas personas consideran más detallada o rica la información que le ofrece la versión BOS (Figura 11).

[1:35] En las dos versiones he obtenido resulta..	[1:31] ambos resultados eran bastante parejos,	[1:17] Resulta que en la encuesta normal mi pun..
En las dos versiones he obtenido resultados parecidos,	ambos resultados eran bastante parejos,	Resulta que en la encuesta normal mi puntuación es parecida a la encuesta bar.

**Figura 8.- algunos ejemplos representativos de las contestaciones de los alumnos sobre comparación**

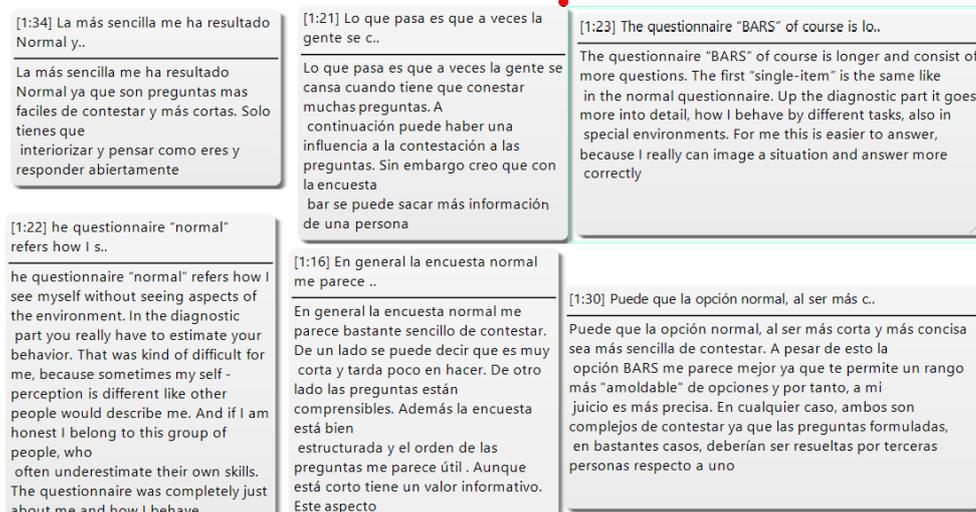
[1:18] Los resultados de la encuesta normal y b..	[1:24] I think, that the questionnaires reflect..	[1:36] Por supuesto los resultados reflejan mi ..	[1:32] Reflejan como CREO que actuó y soy respe..
Los resultados de la encuesta normal y bar creo que reflejan más o menos como soy en referencia a la innovación y al trabajo en equipo. Comparado con mi actitud en trabajos de grupo puedo decir de mi punto de vista que los resultados de las dos encuestas coincide con mi personalidad	I think, that the questionnaires reflect my behavior in a team and how innovative I am, really good, however I am interested how other people thinking about me and if they would answer the same like I did. Of course that must me people which know me good.	Por supuesto los resultados reflejan mi manera de ser aunque en algunos puntos no estaría tan de acuerdo pero en general sí.	Reflejan como CREO que actuó y soy respecto a la innovación y al trabajo en equipo, pero pienso que esta opinión puede diferir si el test lo hiciese sobre mi otra persona, pudiendo tanto aumentar como disminuir

**Figura 9.- algunos ejemplos representativos de las contestaciones de los alumnos sobre acierto del diagnóstico**

[1:20] La encuesta me ha animado pensar sobre m..	[1:26] In my opinion it was meaningful to fill ..	[1:19] Me parece muy útil e interesante para mi..
La encuesta me ha animado pensar sobre mis actividades en grupos. Así podía reflejar mis competencias y me he dado cuenta de unos puntos positivos y debiles.	In my opinion it was meaningful to fill out the questionnaires, because it gives you the possibility to think about yourself and how behave in an environment. However I think it would be good, if other people fill the same questionnaire about me and then it would be interesting if my self-perception is the same like other people think about me.	Me parece muy útil e interesante para mi misma rellenar la encuesta. Todos hemos trabajado varias veces en grupos sin darme realmente cuenta de mi actitud.
[1:27] me ha resultado útil rellenar los cuesti..	[1:14] Por eso creo que los cuestiona..	[1:33] Siempre es útil rellenar cuestionarios d..
me ha resultado útil rellenar los cuestionarios, porque de esa forma puedo ver cual es mi comportamiento a la hora de trabajar en equipo y, a la misma vez, motivarme y mejorar en los aspectos que flojee un poco más, para finalmente conseguir tener un buen comportamiento de interacción, iniciativa, innovación y compañerismo	Por eso creo que los cuestionarios son muy útil para evaluar su mismo para ver que hay una área donde puedo mejorarme, pero no para mandos que quieren saber que persona tiene que mejorar	Siempre es útil rellenar cuestionarios de este tipo para ver que visión se tiene de uno mismo, aunque insisto en que vería más útil aún esta tarea si otras personas con las que he formado equipo lo rellenasen por mí, para contrastar datos y ver en qué puedo estar equivocado respecto a mi manera de ver las cosas y así poder mejorar. Aun que ello conlleva mucho más trabajo

**Figura 10.- algunos ejemplos representativos de las contestaciones de los alumnos sobre utilidad**

*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*



**Figura 11.- algunos ejemplos representativos de las contestaciones de los alumnos sobre comparación**

## Conclusiones

Hemos presentado un modelo para evaluar la competencia transversal de innovación (según el modelo UPV también aparecerían representadas las competencias de creatividad, liderazgo, pensamiento crítico y trabajo en equipo) con estudiantes universitarios en asignaturas de máster con pocos alumnos. El modelo ha demostrado ser válido, con una carga de trabajo razonable y asumible por alumnos, sin excesivas complejidades a la hora de implantarlo y proporcionando unos resultados fiables y muy útiles para los alumnos.

Consideramos que las principales contribuciones de nuestro trabajo son:

- Comparar la fiabilidad de las puntuaciones basadas en ratings y en BOS, así como su viabilidad en un contexto poco trabajado (la docencia universitaria).
- Para los docentes universitarios, analizamos las ventajas e inconvenientes de la aplicación de ambos métodos de toma de datos, para resolver el problema de la evaluación de competencias o habilidades de los alumnos que tienen que ver con el proceso de grupo y no con los resultados o productos del grupo.

## Limitaciones e investigación futura:

La principal limitación es que hemos analizado un grupo de 16 alumnos de máster, lo que nos impide generalizar estos resultados a grupos numerosos o de alumnos de grado. También hemos considerado el mismo peso para los diferentes ítems que componen las escalas o dimensiones de las competencias observadas, de modo que las dimensiones con más criterios, pesan más en la nota ponderada.

Para investigación futura sería interesante ampliar la muestra a más alumnos de master y grado, de diferentes titulaciones y comprobar si, por un lado las opiniones unánimes en la muestra usada en esta investigación, se siguen manteniendo en otras muestras. Además habría que estimar la proporción de alumnos que prefieren cada tipo de autodiagnóstico (pues aparecen diferentes opiniones en nuestra investigación). También debería comprobarse que la supuesta utilidad de opiniones de terceras personas que conocen a los alumnos se confirma cuando realmente sean evaluados por conocidos.

Otra línea de trabajo que consideremos interesante sería hacer un análisis entre los ítems equivalentes en ambos cuestionarios. Los 25 ítems de la versión escala de puntuación están incluidos en la versión BOS pero los niveles de respuesta son frecuencia de manifestación de comportamiento en lugar de grado de excelencia en el comportamiento. Podría comprobarse el efecto de la forma de respuesta en la media y distribución de las respuestas.

## 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 PIME/2014/A/013/A-Barómetro INCODE: Evaluación de Competencias de Innovación en la Empresa y en la Universidad.

## Referencias

- ANDREU ANDRÉS, M. A. & GARCÍA-CASAS, M. (2014) La evaluación de la participación en equipos de trabajo universitarios (Assessment of participation in higher education team working activities). *WPOM-Working Papers on Operations Management*, 5.
- BRINGER, J. D., JOHNSTON, L. H. & BRACKENRIDGE, C. H. (2006) Using Computer-Assisted Qualitative Data Analysis Software to Develop a Grounded Theory Project. *Field Methods*, 18, pp. 245-266.
- DE BEUCKELAER, A., TOONEN, S. & DAVIDOV, E. (2013) On the optimal number of scale points in graded paired comparisons. *Quality and Quantity*, 47, pp. 2869-2882.
- DOWDY, E., TWYFORD, J. & SHARKEY, J. D. (2013) Methods of Assessing Behavior: Observations and Rating Scales. IN D. H. SAKLOFSKE, V. L. SCHWEAN & C. R. REYNOLDS (Eds.) *The Oxford Handbook of Child Psychological Assessment*. (Oxford University Press).
- FRIESE, S. (2012) *Qualitative Data Analysis with ATLAS.ti*, (London, SAGE Publications Ltd).
- GROTH-MARNAT, G. (2003) *Handbook of psychological assessment*, (Hoboken, New Jersey, John Wiley & Sons, Inc.).
- HAIR, J. F., ANDERSON, R. E., TATHAM, R. L. & BLACK, W. C. (1999) *Análisis de datos multivariante [Multivariate data analysis]*, (Madrid, Prentice Hall).
- HATZINGER, R. & DITTRICH, R. (2012) Prefmod: An R package for modeling preferences based on paired comparisons, rankings, or ratings. *Journal of Statistical Software*, 48.
- INGOLS, C. & SHAPIRO, M. (2014) Concrete Steps for Assessing the Soft Skills in an MBA Program. *Journal of Management Education*, 38, pp. 412-435.

*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*

LIND, D. S., REKKAS, S., BUI, V., LAM, T., BEIERLE, E. & COPELAND, I. I. I. (2002) Competency-Based Student Self-Assessment on a Surgery Rotation. *Journal of Surgical Research*, 105, pp. 31-34.

LJUNGMAN, A. G. & SILÉN, C. (2008) Examination involving students as peer examiners. *Assessment & Evaluation in Higher Education*, 33, pp. 289-300.

LOHMANN, A. & PRUMPER, J. (2006) Questionnaire for direct participation in the office (FdP-B) - results concerning its reliability and validity. *Zeitschrift für Arbeits- und Organisationspsychologie*, 50, pp. 119-134.

MARIN-GARCIA, J. A., AZNAR-MAS, L. E. & GONZÁLEZ-LADRÓN-DE-GEVARA, F. (2011) Innovation types and talent management for innovation. *Working Papers on Operations Management*, 2, pp. 25-31.

MARIN-GARCIA, J. A., BAUTISTA POVEDA, Y., GARCIA-SABATER, J. J. & VIDAL CARREAS, P. I. (2010) Implantación de la innovación continua en la gestión de operaciones: una revisión de la literatura. *Innovar*, 20, pp. 77-95.

MARIN-GARCIA, J. A., GARCIA-SABATER, J. J., MAHEUT, J., VALERO-HERRERO, M. & ANDRES-ROMANO, C. (2012) *Gestión de recursos humanos para ingenieros de la rama industrial*, (Harlow, Pearson Education).

MARIN-GARCIA, J. A., GARCIA-SABATER, J. P., MIRALLES, C. & RODRÍGUEZ VILLALOBOS, A. (2008) Profile and competences of Spanish industrial engineers in the European Higher Education Area (EHEA). *Journal of Industrial Engineering and Management*, 1, pp. 269-284.

MARIN-GARCIA, J. A., PEREZ-PEÑALVER, M. J. & WATTS, F. (2013) How to assess innovation competence in services: The case of university students. *Dirección y Organización*, pp. 48-62.

POND, K. (2007) Student Experiences of Peer Review Marking of Team Projects. *International Journal of Management Education*, 6, pp. 30-43.

VAN OVERVELD, K. & VERHOEFF, T. (2013) Self-consistent peer ranking for assessing student work: Dealing with large populations. *CSEDU 2013 - Proceedings of the 5th International Conference on Computer Supported Education*.

WATTS, F., GARCIA-CARBONELL, A. & ANDREU ANDRÉS, M. A. (2013) *Innovation competencies development: INCODE barometer and use guide*, (Turku, Turku University of Applied Sciences).

WATTS, F., MARIN-GARCIA, J. A., GARCIA-CARBONELL, A. & AZNAR-MAS, L. E. (2012) Validation of a rubric to assess innovation competence. *Working Papers on Operations Management*, 3, pp. 61-70.

YIU, E., CHAN, K. & MOK, R. (2007) Reliability and confidence in using a paired comparison paradigm in perceptual voice quality evaluation. *Clinical Linguistics and Phonetics*, 21, pp. 129-145.