

¿Qué sabemos sobre el uso de rúbricas en la evaluación de asignaturas universitarias?

Juan A. Marin-Garcia¹, Cristina Santandreu-Mascarell²

¹ROGLE- DOE- Universitat Politècnica de Valencia (Spain)

²DOE- Universitat Politècnica de Valencia (Spain)

jamarin@omp.upv.es, crisanma@omp.upv.es

Received July, 2014

Accepted March, 2015

Resumen

Objeto: En este artículo comparamos las estrategias de búsqueda y los resultados de tres revisiones de literatura recientes sobre uso rúbricas en educación. Con todo ello pretendemos identificar qué se sabe, que no se sabe aún y si hay material para lanzar una nueva revisión de literatura en estos momentos.

Diseño/metodología/enfoque: Análisis comparado de las tres revisiones de literatura reciente y búsqueda sistemática de literatura.

Aportaciones y resultados: De los 142 artículos incluidos en las tres revisiones recientes, pocos se centran en docencia universitaria y los resultados no se presentan desagregados por niveles educativos. Consideramos que no existe una respuesta científica concluyente para cómo crear y validar una rúbrica para evaluación de asignaturas universitarias, ni para qué sirven las rúbricas, ni si vale la pena el esfuerzo de desarrollarlas. Las recomendaciones para el uso adecuado de rúbricas están basadas en evidencias anecdóticas sin una metodología robusta y contrastada. Tampoco es fácil localizar qué asignaturas, ni qué objetos, ni en qué contextos han sido objeto de investigación las rúbricas, ni qué rúbricas validadas hay disponibles para ellos.

Originalidad / Valor añadido: Hemos realizado una nueva revisión y comparamos los artículos seleccionados con los utilizados en las revisiones anteriores. Hemos localizado

241 nuevos artículos no incluidos en las revisiones anteriores. Probablemente con todo este nuevo material se pueda realizar en el futuro algún meta-análisis.

Palabras clave: Rúbricas, Puntuación, Educación universitaria, Evaluación, Auto-evaluación, Evaluación de pares, Revisión sistemática de literatura

Códigos JEL: A22, I23

Title: What do we know about rubrics used in higher education?

Abstract

Purpose: In this paper we compare the search strategies and the results of three recent literature reviews on the rubrics use in education. With all this we aim to identify what is known, what it is not known yet and if there is material to launch a new literature review right now.

Design/methodology: Comparative analysis of three recent literature reviews and systematic literature review.

Findings: Of the 142 papers included in the three recent reviews, few focus on university learning and the results are not broken down by educational level. We believe that there is no conclusive scientific answer to how to create and validate an assessment rubric for university courses, or what purpose they serve, or if it's worth the effort to develop them. Recommendations for the appropriate use of rubrics are based on anecdotal evidence without robust and proven methodology. Nor is easy to find what subjects, or what objects, and in what contexts have been investigated the rubrics or which validated rubrics are available.

Originality/value: We conducted a new review and we compare the selected papers with those used in previous reviews. We have located 142 new papers not included in previous reviews. Probably all this new material can be used in the future for a meta-analysis.

Keywords: Rating, Scoring, Rubric, Higher Education, Systematic Literature Review, Assessment instrument, Educational Learning, Self-assessment, Peer-assessment

Jel Codes: A22, I23

1. Introducción

El concepto de rúbrica que utilizaremos a lo largo de este artículo está resumido en la siguiente definición (Jonsson & Svingby, 2007; Reddy & Andrade, 2010):

- Una rúbrica es un documento compuesto por tres elementos. Por un lado, describe los criterios que van a ser tenidos en cuenta para la evaluación de un trabajo, tarea, producto o actividad de las personas evaluadas. Por otro, enumera los niveles de calidad de cada uno de los criterios, proporcionando ejemplos de habilidades o características que se deben demostrar para ser asignado un determinado nivel de calidad para un criterio. Por último, en el caso de usarse para evaluación sumativa, incluye los pesos de cada criterio y los puntos asignados a cada nivel de calidad.

El número de artículos científicos publicados, cuya pregunta de investigación está relacionada con las rúbricas, es elevado, teniendo una mayor frecuencia a partir de 1997 (Jonsson & Svingby, 2007). Existen ejemplos de uso de las rúbricas para evaluar diferentes actividades de alumnos universitarios en asignaturas de gestión (management) tanto en estudios de Ingeniería (Andreu Andrés y García-Casas, 2014; Marín-García, 2009; Watts, Marín-García, García-Carbonell & Aznar-Mas, 2012) como en Administración de Empresas (McGoldrick & Peterson, 2013).

De hecho, hemos localizado tres revisiones de literatura publicadas recientemente en revistas incluidas en las bases de datos de EBSCO, Web Of Knowledge o Science Direct (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013; Reddy & Andrade, 2010). En este artículo vamos a apoyarnos en estas revisiones previas para intentar dar respuesta a las siguientes cuestiones relacionadas con el uso de rúbricas en educación universitaria.

- Como crear y validar una rúbrica para evaluación de asignaturas universitarias
- ¿Para qué sirven las rúbricas (funciones, ventajas, inconvenientes)? ¿vale la pena el esfuerzo de desarrollarlas?
- ¿Qué recomendaciones se dan para el uso adecuado de rúbricas?
- ¿Qué asignaturas, qué objetos y en qué contextos han sido objeto de investigación las rúbricas y qué rúbricas validadas hay disponibles para ellos?
- ¿Qué es necesario investigar en el futuro para seguir avanzando en el conocimiento científico sobre rúbricas en enseñanza universitaria?
- ¿Hay material nuevo, no incluido en las tres revisiones citadas, que aconseje hacer una nueva revisión de literatura?

Nuestro trabajo extiende las revisiones previas documentando el proceso de revisión de modo que se puede replicar y ampliar nuestro trabajo y aportando un conjunto de referencias nuevas publicadas en el periodo 2000-2013 y proponiendo una estrategia de búsqueda adecuada para una revisión sistemática de literatura que se ajuste a las preguntas de investigación planteadas. Estas nuevas referencias no son sólo debidas a incorporar publicaciones posteriores a la fecha de realización de las revisiones publicadas, sino que incluyen nuevas referencias que fueron falsos negativos en los trabajos anteriores o que no encajaban en los criterios de inclusión de las revisiones realizadas hasta la fecha.

2. Conclusiones de las revisiones de literatura recientes

Empezaremos este apartado resumiendo las principales características de las tres revisiones de literatura recientes (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013; Reddy & Andrade, 2010). En la Tabla 1 podemos comprobar que las preguntas de investigación de las tres revisiones difieren bastante y esto originará que el conjunto de artículos analizados en cada una de ellas sea diferente y, del mismo modo, el enfoque del análisis descriptivo realizado se centrará en aspectos distintos. La revisión de Reddy y Andrade (2010) se centra en educación universitaria y es la más exploratoria de las tres y se centra en cuestiones generales intentando ver qué se ha investigado y cuáles son las posibles líneas de investigación futura. Una de sus preguntas, si las rúbricas son capaces de promover un mayor aprendizaje en los alumnos, es compartida por Jonsson & Svingby (2007) que además, se centran en analizar la capacidad de las rúbricas en mejorar la fiabilidad o validez de las puntuaciones, sin diferenciar nivel educativo. Por último, Panadero y Jonsson (2013) se centran exclusivamente en el uso de rúbricas para la evaluación formativa, siendo los más cualitativos en la presentación de resultados al renunciar explícitamente a intentar cuantificar la fortaleza o significación de los efectos de usar rúbricas en contextos de evaluación formativa.

Las tres revisiones utilizan diferentes bases de datos. Las tres incluyen Educational Resources Information Center (ERIC) y PsychINFO, pero sólo dos de ellas (Jonsson & Svingby, 2007; Reddy & Andrade, 2010), incluyen las revistas de Science Direct y las bases de datos ABI o EBSCO (que tienen un perfil un poco diferente de revistas, al compararla con la bases de datos de Web Of Knowledge y Science Direct). Ninguna de las revisiones hace explícitos los años en los que se ha acotado la búsqueda. Dos de ellas (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013) indican que no han establecido límites, lo que significa que la fecha de inicio de búsqueda es cualquier año, pero queda indeterminado hasta qué año se ha incluido. No obstante, analizando los artículos seleccionados, los más antiguos son de 1995 y los más recientes entre 2006 y 2011 dependiendo de la revisión. En las tres revisiones analizadas, las

estrategias de búsqueda no están claramente especificadas. Se comentan las palabras clave que se han usado pero no se hace mención a los campos en los que las han buscado ni la combinación de operadores (AND, OR, Etc) usados exactamente. Tampoco se concreta si son las únicas palabras clave empleadas o si son sólo un ejemplo de algunas de las usadas, lo que hace complicada la replicación de dichas búsquedas. Sin embargo, todas han detallado con claridad los criterios de inclusión/exclusión. Analizando estos criterios, todas ellas se basan en investigaciones empíricas sobre el uso de rúbricas. Sólo la investigación de Reddy y Andrade (2010) se centra en educación universitaria y admite artículos sobre desarrollo de rúbricas, validación de las mismas o ventajas e inconvenientes de uso de las rúbricas. Las otras dos revisiones usan artículos centrados en cualquier nivel educativos pero excluyendo el desarrollo y validación de rúbricas, centrándose sólo en artículos que utilizan rúbricas para evaluar el rendimiento académico en general (Jonsson & Svingby, 2007) o sólo la evaluación formativa (Panadero & Jonsson, 2013).

	(Reddy & Andrade, 2010)	(Jonsson & Svingby, 2007)	(Panadero & Jonsson, 2013)
Pregunta de investigación	<p>¿Qué tipo de investigación se ha realizado sobre las rúbricas en la educación superior?</p> <p>¿Hay evidencia de que las rúbricas puedan ser utilizadas como evaluaciones formativas con el fin de promover el aprendizaje y el logro de la educación superior, en comparación con el uso de las rúbricas que sólo sirven a los fines de clasificación y la obtención de resultados?</p> <p>¿Cuánta atención se ha prestado a la calidad de las rúbricas utilizadas por colegios y universidades instructores?</p> <p>¿Cuáles son algunas de las directrices que podrían ser fructíferas para la investigación y el desarrollo futuro de esta área?</p>	<p>¿El uso de las rúbricas ha mejorado la fiabilidad de la calificación?</p> <p>¿Pueden las rúbricas facilitar la determinación válida de las evaluaciones del rendimiento?</p> <p>¿El uso de las rúbricas puede promover el aprendizaje y / o mejorar la enseñanza?</p>	<p>Para identificar un espectro/variedad/gama de los efectos potenciales del uso de las rúbricas formativas, así como de los diferentes factores, que pueden moderar estos efectos.</p>
Bases de datos	<p>ABI Inform Global, Academic Search Premier, Blackwell Journals, CBCA Education, Sage Journals Online, Emerald Full Text, SpringerLink, ProQuest Education Journals, ERIC, J-Store, PsychInfo, Education Research Complete and EBSCO.</p>	<p>Educational Resources Information Center (ERIC), PsychINFO, Web of Science, ScienceDirect, Academic Search Elite/EBSCO, JSTOR and Blackwell Synergy.</p>	<p>PsycINFO y ERIC databases</p>

	(Reddy & Andrade, 2010)	(Jonsson & Svingby, 2007)	(Panadero & Jonsson, 2013)
Periodo años	No especificado. Los artículos seleccionados son de 2001 a 2007	Fecha de inicio cualquier año No especificada fecha de fin Los artículos seleccionados son de 1995-2006	Fecha de inicio cualquier año No especificada fecha de fin. Los artículos seleccionados son de 2001 a 2011
Estrategia de búsqueda	Las palabras clave buscadas incluyen rúbrica, la educación superior, post-secundaria y estudios empíricos.	rubricas/rubricas educativas rubricas/rubricas de valoración/puntuación	La rúbrica/s y la autoevaluación; la rúbrica/s y la evaluación en grupo; la rúbrica/s y el aprendizaje autorregulado/ autorregulación. Añadiendo referencias con el método ‘snowballing method’
Criterios de Inclusión/ exclusión	‘Investigación empírica’ y la ‘Educación Superior’. Se excluyen las tesis de master. Las tesis doctorales se incluyeron si parecían utilizar métodos de investigación que podrían llevar a resultados creíbles	Sólo los artículos revisados por pares en revistas, comunicaciones a congresos, informes de investigación y tesis doctorales. Sólo los estudios que informan de manera explícita en la investigación empírica donde se utilizaron las rúbricas para la evaluación del rendimiento excepto los artículos sobre el desarrollo de las rúbricas, artículos de opinión sobre los beneficios de las rúbricas, y guías sobre el uso de rúbricas.	Sólo el material impreso y revisado por pares, como artículos en revistas, libros editados, informes de investigación y tesis doctorales. Si se basa en datos empíricos sobre el uso de rúbricas. El uso de rúbricas tenía que ser con fines formativos. Los argumentos teóricos o artículos que presentan sólo directrices prácticas no fueron incluidos en la revisión.

Tabla 1. Resumen de las revisiones de literatura recientes

El conjunto de artículos analizados en las tres revisiones suma 142 artículos diferentes (Tabla 2), 25 usados exclusivamente en Reddy y Andrade (2010); 59 sólo en Jonsson y Svingby (2007) y 37 sólo en Panadero y Jonsson (2013). El resto de artículos aparecen sólo en dos de las revisiones o en las tres.

Referencia	Número de artículos en cada combinación
R-J-P	7
R-J	3
R-P	2
J-P	9
R	25
J	59
P	37
Total de referencias	142

Tabla 2. Resultados de las revisiones de literatura reciente. Cada artículo se ha contado sólo una vez clasificado en la categoría correspondiente. R: Reddy y Andrade (2010); J: Jonsson y Svingby (2007); P: Panadero y Jonsson (2013)

A continuación resumiremos la información contenida en estas revisiones relacionada con nuestras preguntas de investigación. No obstante debemos advertir que todos los resultados de las revisiones provienen de una interpretación basada en “vote counting” y “narrative review”, lo que impide extraer conclusiones definitivas (Borenstein, Hedges, Higgins & Rothstein, 2013). Por ello, entendemos que todas las respuestas que se exponen a continuación deben considerarse como proposiciones tentativas. Estas proposiciones no han sido validadas adecuadamente por falta de evidencia científica publicada que permita generalizar los resultados de investigaciones heterogéneas que no han sido suficientemente replicadas.

2.1. Como crear y validar una rúbrica para evaluación de asignaturas universitarias

Debido al objetivo de las revisiones analizadas y sus criterios de inclusión/exclusión, ninguna de ellas aporta información sobre cómo crear una rúbrica. Pero aportan cierta información que nos permite intuir parte del proceso a seguir para comprobar su calidad como instrumento de medida, ya que analizan aspectos como la validez (si se mide lo que se intenta medir) y la fiabilidad (consistencia) de las rúbricas.

El número de referencias que analizan explícitamente la validez de las rúbricas es muy escaso (Reddy & Andrade, 2010). No obstante, podemos considerar que el primer paso es que la rúbrica esté escrita en un lenguaje claro que no dé pie a interpretaciones ambiguas (Reddy & Andrade, 2010). A partir de ahí hay que demostrar la relevancia y representatividad del contenido de la rúbrica (Doval Diéguez & Viladrich Segués, 2011; Jonsson & Svingby, 2007).

Es decir, que no falta ni sobra ninguno de los criterios para representar los aspectos importantes de lo que se pretende evaluar. Para justificar la validez es necesario hacer una revisión de los instrumentos similares y/o contar con la opinión de expertos. También se puede realizar análisis estadísticos para ver la consistencia interna de los ítems o dimensiones que van a resumirse en un solo valor, o en qué medida las puntuaciones obtenidas con la rúbrica convergen con otras medidas ya validadas para evaluar aprendizajes similares o discriminan las puntuaciones de los estudiantes clasificados por otros procedimientos o concuerdan con otras medidas que la teoría haya demostrado que deben ser consecuencia o estar muy asociadas a los aprendizajes buscados (Delgado Rodríguez, Llorca & Doménech Massons, 2012; Viladrich Segués y Doval Diéguez, 2011).

Por otra parte, la fiabilidad nos indica si las puntuaciones obtenidas con la rúbrica son consistentes y es una condición necesaria, pero no suficiente, para la validez (Reddy & Andrade, 2010). Los índices de fiabilidad nos indican si las puntuaciones son similares cuando evaluamos situaciones parecidas (Delgado Rodríguez et al., 2012; Hair, Anderson, Tatham & Black, 1999; Reddy & Andrade, 2010). Para evaluar la fiabilidad tenemos dos alternativas. Podemos medir el consenso/agreement (fiabilidad absoluta, medida como grado en que diferentes evaluaciones proporcionan la misma puntuación). O podemos medir la consistencia (fiabilidad relativa basada en medida de la asociación entre las puntuaciones) (Jonsson & Svingby, 2007; Reddy & Andrade, 2010). La consistencia (o fiabilidad relativa) puede ser intra-rater (comparar las puntuaciones de un mismo evaluador en momentos diferentes, usando los mismos métodos o instrumentos –situaciones, productos, rubricas...-, estando los evaluados en las mismas condiciones –su nivel de aprendizaje no ha cambiado-); inter-rater (comparar las puntuaciones de diferentes evaluadores que puntúan simultáneamente, utilizando los mismos métodos y procedimientos) (Hair et al., 1999; Jonsson y Svingby, 2007). Adicionalmente, se pueden utilizar métodos para estimar el error relativo y absoluto de la varianza de las puntuaciones de los evaluadores utilizando los coeficientes de “Dependability y Generalizability” de la “Generalizability theory” o los valores derivados de la F del análisis ANOVA (Jonsson & Svingby, 2007; Marín-García, Miralles Insa & Marín García, 2008; Reddy & Andrade, 2010; Viladrich Segués & Doval Diéguez, 2011). Probablemente, para la validación de rúbricas, la medida de consistencia más crítica y que, por lo tanto debe garantizarse en todas las ocasiones, sea la fiabilidad intra-rater (Jonsson & Svingby, 2007).

Existen diferentes índices con los que medir la fiabilidad (Delgado Rodríguez et al., 2012; Jonsson & Svingby, 2007; Reddy & Andrade, 2010). En la Tabla 3 resumimos los estadísticos informados en las revisiones sistemáticas sobre rúbricas. Para el grado de Consenso/Agreement se puede usar el acuerdo total entre evaluadores o el índice de acuerdo más/menos un punto de la escala. Ambas medidas tienen el inconveniente de que dependen

en gran medida del número de niveles de la escala y que no tienen en cuenta la coincidencia esperable sólo por azar. Todo ello las hace poco útiles para comparar diferentes rúbricas entre sí. Por este motivo, se suele usar el estadístico Cohen's kappa, cuyos niveles de referencia (Tabla 1) sólo se pueden aplicar en caso de analizar datos dicotómicos, puesto que al aumentar el número de posibles categorías de respuesta se modificarían estos valores. Probablemente, el coeficiente de correlación interclase de acuerdo, bien con efectos fijos o efectos aleatorios, o el gráfico de Bland-Altman o el método de Kaplan-Meier, serían una medida más adecuada en los casos en que las varianzas de las puntuaciones sean similares (Delgado Rodríguez et al., 2012). Los porcentajes de acuerdo informados por Jonsson y Svingby (2007) no se acompañan de información sobre la cantidad de niveles de puntuación de las rúbricas ni de los acuerdos específicos en positivo y negativo (Delgado Rodríguez et al., 2012) que serían necesarios para interpretar esos valores. Los valores de Kappa podrían indicar una fiabilidad entre moderada y buena, sin embargo, vuelve a escasear la información adicional necesaria y, en cualquier caso, el número de estudios disponibles es excesivamente bajo dada la presumible heterogeneidad de los efectos. Para medir la fiabilidad intra-rater e inter-rater, se han usado habitualmente los índices de correlación que, aún siendo una medida fácil de calcular, no deberían ser usados para medir la reproducibilidad de puntuaciones con variables continuas. En su lugar, deberían usarse pruebas t-student para muestras emparejadas o alguno de los coeficientes de correlación interclase (ICC) puesto que permiten analizar tanto el grado de acuerdo como la consistencia de las puntuaciones (Delgado Rodríguez et al., 2012). Sólo una de las revisiones informa acerca del grado de consistencia de las puntuaciones (Jonsson & Svingby, 2007) obteniéndose valores aceptables de correlación en las puntuaciones inter-rater. En ningún caso se han usado estadísticos ICC. Por otra parte, al analizar el alfa de Cronbach, no se especifica explícitamente si los análisis de los 8 estudios analizados son para consistencia inter-rater o intra-rater. En cualquier caso, el número de estudios vuelve a ser muy limitado.

La fiabilidad no es tan importante en usos formativos como con evaluación sumativa, pues las consecuencias y la dinámica son distintas. Cuando se usan las rúbricas para evaluación formativa, hay una interacción frecuente, incluso un dialogo, entre profesor y alumno que permite matizar o corregir las percepciones de la evaluación y ajustarlas a la verdadera situación del estudiante y poder programar acciones para mejorar habilidades o competencias concretas. Por ello, unos niveles bajos de fiabilidad no son excesivamente problemáticos, ya que hay posibilidad de ajustar la puntuación posteriormente. Sin embargo, en la evaluación sumativa es necesario garantizar, por una lado, una buena fiabilidad intra-evaluador y, por otro lado, si intervienen varios profesores puntuando de manera independiente el trabajo de alumnos, además es imprescindible garantizar una adecuada fiabilidad inter-evaluadores, para garantizar que todos los alumnos han sido puntuados de manera similar (Jonsson & Svingby, 2007).

	Nivel de referencia	(Reddy & Andrade, 2010)	(Jonsson & Svingby, 2007)	(Panadero & Jonsson, 2013)
% de acuerdo total entre evaluadores	Depende del número de niveles de la escala	n.a.	55-75% (N=18)	n.a.
% acuerdo más menos un punto de escala (adjacent agreement)	Depende del número de niveles de la escala	n.a.	Cercano al 90% (N=14)	n.a.
Cohen's kappa Inter-rater	0.61 o por encima buena o muy Buena 0.40-0.60 moderado (Delgado Rodríguez et al., 2012; Jonsson & Svingby, 2007; Reddy & Andrade, 2010)	0.41 (n=1)	0.20-0.63 (pocos por debajo de 0,40) (N=4)	n.a.
Pearson's o Spearman's Ro correlation Inter-rater	<0.5 moderado bajo >0.7 alto (Delgado Rodríguez et al., 2012; Hair et al., 1999)	n.a.	0.27-0.98 con la mayoría entre 0.55-0.75 (n=4) (n=9)	n.a.
Cronbach's alpha Inter-rater or intra-rater	0.70 (Delgado Rodríguez et al., 2012; Hair et al., 1999)	n.a.	0.50-0.92 con la mayoría por encima de 0.70 (n=8)	n.a.
Generalizability theory	0.80 (Jonsson & Svingby, 2007)	Dependability 0.71 Generalizability 0.65 (n=1)	Dependability 0.06-0.96 Generalizability 0.15-0.98 (n=15)	n.a.
ANOVA intraclass	n.a.	n.a.	?? (n=1)	n.a.

N es el número de estudios. n.a.= no disponible

Tabla 3. Resultados de las revisiones de literatura reciente

El número de estudios que analizan la consistencia de las puntuaciones usando rúbricas son muy escasos e impiden hacer interpretaciones concluyentes. Esta escasez es todavía mayor en lo que se refiere a la fiabilidad intra-evaluador (Jonsson & Svingby, 2007).

Cuando los estudiantes pueden realizar diferentes tareas o elegir temas distintos para sus trabajos o generar productos únicos para ser evaluados, como por ejemplo proyectos, presentaciones orales, ensayos o similares, es muy difícil conseguir altos índices de fiabilidad, por lo que los niveles de fiabilidad, que en estadística tradicional se consideran como adecuados, quizás deban ser más bajos cuando estemos tratando rúbricas diseñadas para evaluar tareas complejas (Jonsson & Svingby, 2007). Por otra parte, independientemente del grado de fiabilidad que se consiga con una rúbrica, quizás lo que interese es saber si el hecho de usar o no usar una rúbrica permite mejorar el nivel de partida de fiabilidad que se tiene a la

hora de evaluar una actividad en un contexto docente determinado y si se puede hacer algunas cosas para lograr mejorar la fiabilidad de las rúbricas usadas. Ambos aspectos serán analizados en las secciones siguientes.

2.2. ¿Para qué sirven las rúbricas? ¿Vale la pena el esfuerzo de desarrollarlas?

Ventajas	Causa	Evidencia
Los estudiantes aprenden más cuando son evaluados con rúbricas (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013; Reddy & Andrade, 2010)	Las expectativas y criterios de los profesores son más explícitos (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013) Ayuda a los estudiantes a comprender los objetivos y clarificar qué se va a puntuar y cómo (Panadero & Jonsson, 2013; Reddy & Andrade, 2010) Los estudiantes pueden evaluar mejor su trabajo (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013; Reddy & Andrade, 2010) Las puntuaciones son más transparentes (Panadero & Jonsson, 2013; Reddy & Andrade, 2010) Los estudiantes pueden regular mejor su progreso (Panadero & Jonsson, 2013; Reddy & Andrade, 2010) Mejoran la motivación del alumno (Jonsson & Svingby, 2007) Mejoran el pensamiento reflexivo y crítico de los alumnos (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013) Se reduce el nivel de ansiedad de los estudiantes, si hay suficiente tiempo para cumplir con las exigencias (Panadero & Jonsson, 2013)	Evidencia muy limitada. Conclusiones basadas en "vote counting" de estudios heterogéneos con muy baja N
Aumenta la credibilidad de las puntuaciones. Los profesores realizan la evaluación de forma más eficiente (Jonsson & Svingby, 2007; Reddy & Andrade, 2010)	Las rúbricas hacen la evaluación más objetiva y justa (Reddy & Andrade, 2010) Las rúbricas ayudan a que las puntuaciones de los profesores sean más fiables (Jonsson & Svingby, 2007; Reddy & Andrade, 2010) Las rúbricas permiten realizar la evaluación invirtiendo menos tiempo (Reddy & Andrade, 2010) Las rúbricas permiten evaluar el rendimiento de competencias que difícilmente se puede evaluar en un examen escrito (Jonsson & Svingby, 2007)	Evidencia muy limitada. Conclusiones basadas en "vote counting" de estudios heterogéneos con muy baja N
Mejora la calidad del feedback que pueden dar los profesores a los alumnos (Reddy & Andrade, 2010) (Jonsson & Svingby, 2007)	Las expectativas y criterios de los profesores son más explícitos (Jonsson & Svingby, 2007) Con rúbricas, los profesores descubren más fácilmente los puntos débiles de los alumnos e identifican las áreas de mejora necesarias en la formación (Jonsson & Svingby, 2007; Reddy & Andrade, 2010)	No concluyente. Impresión obtenida de algún estudio disperso.

Tabla 4. Ventajas del uso de rúbricas en educación

Inconveniente	Causa	Evidencia científica
Resistencia por parte de los profesores a usar rúbricas (Reddy & Andrade, 2010)	Poca preparación en nuevas tendencias de evaluación y desconocimiento del propósito y características de las rúbricas (Reddy & Andrade, 2010)	No concluyente. No parece haber ningún estudio cuyo principal objetivo sea cuantificar el grado de resistencia de los profesores hacia el uso de rúbricas y las causas de esa resistencia
El uso de rúbricas requiere tiempo y esfuerzo	El proceso de creación o adaptación, así como su validación adecuada es muy costoso y complejo (Reddy & Andrade, 2010) Probablemente el uso de rúbricas suponga también un coste de tiempo para alumnos y profesores (no explicitado en las tres revisiones analizadas)	Evidencia anecdótica y no informada científicamente del costo real (para alumnos y profesores) de usar rúbricas comparado con otras formas de evaluación que generen resultados similares.
Las rúbricas por si solas no mejoran el rendimiento de los alumnos (Panadero & Jonsson, 2013; Reddy & Andrade, 2010)	Es necesario que los alumnos se formen en su uso, las comprendan, las usen para su autoevaluación (Reddy & Andrade, 2010) Hay que utilizarlas en un contexto adecuado (Auto-evaluación del alumno o actividades de meta cognición) (Panadero & Jonsson, 2013)	No concluyente. Muy pocos estudios enfocados a responder esta cuestión
Las rúbricas sin “anclas” (descripciones o ejemplos que muestran los diferentes niveles de rendimiento) no son útiles (Jonsson & Svingby, 2007)	Los alumnos tienen problemas para interiorizar los criterios de calidad y trasladarlos a una puntuación numérica que sea comparable a la de los profesores (Jonsson & Svingby, 2007)	Poca información en la literatura publicada para identificar/cuantificar la aportación o mejora de las anclas al uso de rúbricas, percepciones de los alumnos o evaluadores y fiabilidad o validez de las puntuaciones
Los estudiantes hacen sólo lo que se va a medir en la rúbrica (Panadero & Jonsson, 2013)	Las rúbricas hacen que los estudiantes se centren en lo superficial (estrictamente lo que se va a puntuar) y no en el aprendizaje profundo (Panadero & Jonsson, 2013)	No concluyente. Impresión obtenida de algún estudio disperso.

Tabla 5. Inconvenientes del uso de rúbricas en educación

En las Tablas 4 y 5 resumimos las principales ventajas e inconvenientes asignadas al uso de rúbricas en educación universitaria. También se incluyen las causas que promueven esas ventajas e inconvenientes. Estas aportaciones nos permiten dar una respuesta incipiente y tentativa a la primera de las preguntas, ¿para qué sirven las rúbricas? En general todas las conclusiones mostradas deben entenderse como proposiciones pendientes de validación empírica pues el número de investigaciones disponibles es muy escaso, se han realizado con objetivos, métodos y contextos muy diferentes que pueden dar lugar a heterogeneidad de los efectos de las rúbricas y, en la mayoría de los casos, los estudios se han limitado a un tamaño muestra muy pequeño. Todo ello impide generalizar los resultados. Por otra parte, para agregar los resultados, los autores consultados se han valido de una revisión narrativa (“vote

counting”), lo que, en el contexto comentado puede dar lugar a interpretaciones equivocadas (Borenstein et al., 2013).

Para la segunda pregunta del encabezado, ¿vale la pena el esfuerzo de desarrollar rúbricas? Aún no estamos en disposición de poder dar una respuesta. En primer lugar, es necesario validar empíricamente las proposiciones sobre ventajas e inconvenientes. En segundo lugar, es necesario investigar qué esfuerzo en tiempo/recursos requiere el desarrollo de nuevas rúbricas o la adaptación de las creadas por otros docentes. Cuando la investigación haya avanzado para dar respuestas a estos dos temas, podremos valorar si las ventajas derivadas del uso de rúbricas compensan los inconvenientes o el coste del desarrollo o adaptación.

2.3. ¿Qué recomendaciones se dan para el uso adecuado de rúbricas?

En el apartado anterior hemos comentado ventajas e inconvenientes de las rúbricas. Muchos de ellos se darán o no en función del modo en que se hayan usado las rúbricas. En el estado actual de la investigación, las recomendaciones más frecuentes que se hacen para el uso adecuado de rúbricas son las siguientes:

- Que los estudiantes dispongan de la rúbrica desde el principio y las utilicen para realizar una evaluación formativa de su tarea en curso (no sólo para evaluar el producto final una vez entregado) (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013; Reddy & Andrade, 2010).
- Diseñar y validar las rúbricas adecuadamente teniendo en cuenta el contexto y los objetivos de aprendizaje, para asegurar su validez (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013; Reddy & Andrade, 2010).
- Adaptar la rúbrica a las características de los alumnos, tiempo disponible y tarea encargada (Reddy & Andrade, 2010).

Además, las recomendaciones específicas para mejorar la fiabilidad inter-rater:

- Utilizar rúbricas analíticas en lugar de genéricas. Esta recomendación tiene sentido si cada ítem es específico y ha sido demostrada la validez convergente de estos ítems. En este caso, la fiabilidad de la puntuación obtenida al agregar varios ítems es mayor. Sin embargo, si los criterios representan dimensiones independientes del aprendizaje y se trata de competencias genéricas, probablemente sea mejor usar una rúbrica holística que ayude a discriminar los rendimientos excelentes de los pobres (Jonsson & Svingby, 2007; Reddy & Andrade, 2010).

- Reducir el número de niveles de calidad mejora los estadísticos de consenso (las escalas dicotómica serían preferibles en este caso). Sin embargo, tener más niveles ayuda a mejorar la consistencia de las puntuaciones (escalas con 3-5 niveles y subniveles -, =, + en cada uno de ellos) (Jonsson & Svingby, 2007; Reddy & Andrade, 2010).
- Incluir anclas o ejemplos para cada uno de los niveles principales de desempeño (de 2 a 5 niveles). Al elegir las anclas o ejemplos hay que ser muy específico y cuidadosos pues son ellos los que acabarán definiendo cómo interpretar los criterios (Jonsson & Svingby, 2007; Reddy & Andrade, 2010).
- Se debe formar suficientemente a los evaluadores para conseguir mejorar la fiabilidad de las puntuaciones. Cuantos más criterios y más niveles de desempeño se incluyan más formación es necesaria (Jonsson & Svingby, 2007; Reddy & Andrade, 2010).
- Si se utiliza a los alumnos como evaluadores de sus compañeros, el promedio de dos a cuatro puntuaciones de alumnos suelen ser suficientes para lograr unos niveles de fiabilidad parecidos a los de profesores formados para el uso de la rúbrica (Jonsson & Svingby, 2007).

2.4. ¿Qué asignaturas, qué objetos y en qué contextos han sido objeto de investigación las rúbricas y qué rúbricas validadas hay disponibles para ellos?

La mayoría de las rúbricas analizadas se centran en evaluar productos de los estudiantes: memorias de proyectos (revisiones de literatura, portafolios) o ensayos (reflective writings), informes de laboratorio, presentaciones orales, posters, mapas conceptuales (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013; Reddy & Andrade, 2010) y son escasas las rúbricas elaboradas para evaluar los procesos seguidos o las competencias adquiridas por los estudiantes para realizar diversas tareas como pensamiento crítico, trabajo en equipo, gestión de proyectos, etc.

En los últimos años, los artículos publicados parecen centrarse cada vez más en el uso de rúbricas para auto-evaluación o evaluación de pares (Jonsson & Svingby, 2007).

Por otra parte, son muchas las asignaturas o titulaciones en las que se han investigado las rúbricas: biology, management, psychology, leadership, information technology, medicine, nursing, dentistry, food science, teacher education and film technology (Panadero & Jonsson, 2013; Reddy & Andrade, 2010). Sin embargo existe una alarmante carencia de replicación, dada la escasez de artículos, lo variado de los contextos investigados y los pequeños tamaños

muestrales de las investigaciones (en la mayoría de los casos los trabajos se limitan a un grupo de clase y N menor que 100). Este hecho, añadido a la evidencia de que la utilidad, eficiencia o efectos de las rúbricas dependen del contexto donde se usen, hace imposible extraer conclusiones definitivas mientras no haya un conjunto más amplio de estudios que compartan objetivos, métodos y contextos para hacer una adecuada síntesis de los resultados que permitan extraer conclusiones generalizables.

2.5. ¿Qué es necesario investigar en el futuro para seguir avanzando en el conocimiento científico sobre rúbricas en enseñanza universitaria?

En los tres artículos de revisión analizados se comentan diferentes líneas de investigación futura relacionadas con las rúbricas en educación universitaria:

- Utilizar diseños de investigación más robustos para dar respuesta a las preguntas de investigación que se han comentado en los cuatro apartados anteriores de esta investigación. Se sugieren experimentos, cuasi-experimentos o casos y controles con grupos de control y pre-post test, incorporando las variables de confusión adecuadas para aislar sus efectos (por ejemplo, género, estilos de aprendizaje de los alumnos, rendimiento académico previo, horas de dedicación, actividades de meta-cognición como auto-evaluación o evaluación de compañeros, motivación, confianza, etc.) (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013; Reddy & Andrade, 2010).
- Utilizar las técnicas de análisis adecuadas para el diseño y los objetivos e informar de los "effect size" desagregados para la diferentes variables de confusión (tipo de evaluador, tipo de alumno) y los datos detallados de la intervención para poder integrar la información en un meta-análisis futuro (materia, nivel educativo, duración de la intervención, cómo se ha creado/adaptado la rúbrica, cómo se ha validado, que niveles de fiabilidad se han obtenidos, si ha habido formación previa y la duración de esa formación) (Panadero & Jonsson, 2013; Reddy & Andrade, 2010).
- La evidencia científica actual sobre el tema es rudimentaria debido al escaso número de investigaciones. Es preciso replicar los trabajos previos ya que la mayoría se han realizado en los grupos de alumnos de los propios autores. En especial se recomienda replicar en países diferentes de Estados Unidos (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013; Reddy & Andrade, 2010).
- Investigar la relación entre el resultado de uso de rúbricas y las conductas de auto-regulación de los estudiantes (Panadero & Jonsson, 2013; Reddy & Andrade, 2010).

- Comprobar si la auto-evaluación es una tarea más compleja que la evaluación del trabajo de otras personas (Jonsson & Svingby, 2007).

Además de las propuestas de investigación futura contenidas en las tres revisiones, con el fin de poder estimar si hay material suficiente para un meta-análisis, consideramos necesario crear un listado exhaustivo de artículos clasificados por combinación de producto o proceso que evalúan, rúbrica empleada, contexto de uso (datos detallados de la intervención), así como objetivos y efectos medidos en la investigación.

Es posible que la materia no sea una variable que modifique los efectos del uso de las rúbricas (Panadero & Jonsson, 2013). Sin embargo, el nivel de la tarea a evaluar (reproducción de conocimiento, clasificación, etc. frente a alto nivel conceptual como análisis, síntesis, crítica de conocimiento), el modo de usar las rúbricas, el contexto nacional o cultural de la institución universitaria y, quizás, cuestiones de género asociadas a la percepción de auto-eficacia, sí que pueden ser origen de heterogeneidad de los efectos (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013).

Llevar a cabo una única investigación que pueda contemplar todas estas variables y conseguir un tamaño muestral suficiente para generalizar los resultados es una tarea muy complicada. Por lo tanto, sugerimos recoger suficientes estudios que informen adecuadamente de las variables para lograr, en conjunto, un N agregado suficiente para sacar conclusiones generalizables a cada cruce de variables (Marín-García, 2013).

2.6. ¿Hay material nuevo, no incluido en las tres revisiones citadas que aconseje hacer una nueva revisión de literatura?

Teniendo en cuenta que las respuestas proporcionadas por la investigación actual a nuestras preguntas de investigación son insuficientes y que las estrategias de búsqueda de las revisiones anteriores no ha sido informada con claridad, que los criterios de exclusión limitan artículos que podrían ser interesantes para los objetivos de este trabajo y que, dos de las revisiones, no contienen artículos desde hace 5 o más años, es posible que hay nuevo material que aconseje realizar una nueva revisión.

En la Tabla 6 resumimos los resultados de las estrategias de búsqueda realizadas. Los criterios de inclusión/exclusión utilizados para filtrar manualmente las referencias extraídas de la búsqueda automática fueron:

- Incluir cualquier trabajo que se centre en uso de rúbricas en educación universitaria (grado o máster) para evaluar el rendimiento de los alumnos en trabajos académicos o superación de objetivos de aprendizaje de las materias.

- Excluir:
 - Otros niveles educativos (K-12, escuela, educación de adultos)
 - Evaluación de profesionales en empresas u organizaciones
 - Selección de personal
 - Diagnóstico médico o psicológico
 - Evaluación de cuestionarios para diagnóstico organizativo o investigación
 - Idiomas fuera de una titulación universitaria

Las tres primeras estrategias incluyen un filtro específico para detectar sólo referencias con "higher education" o "university" en su título, resumen o palabras clave. El total de referencias encontradas es manejable (131) y con pocos falsos positivos. En total se han obtenido 48 referencias que respetan los criterios de inclusión/exclusión, tras eliminar duplicados y filtrado manual por título y resumen.

Con la finalidad de comprobar si nuestras estrategias dan origen a falsos negativos, hemos retirado el filtro específico de "higher education" Or "universit*" buscando en Science direct. Con ello, hemos pasado de 12 a 387 resultados (casi 20 veces mayor) y tras un filtrado manual, sólo cumplían los criterios 28 referencias (hemos pasado de 4 a 28 referencias). Las nuevas referencias indican el nombre de la materia universitaria pero no hacen mención a "higher education" ni a "university" en su título, resumen o palabras clave. Extrapolando estos resultados a las otras bases de datos, podemos estimar que al retirar el filtro "AND ("higher education" OR Universit*)" el total de referencias obtenidas inicialmente en las tres bases de datos hubiese sido en torno a 4,000, si se mantuvieran las proporciones encontradas en sciencedirect. Por otra parte, dado que la mayor fuente de falsos positivos en esta estrategia es originada por las palabras clave "OR (Assessment instrument)" en el título, hemos mejorado la estrategia de búsqueda usada definitivamente (ver las 4 últimas filas de la Tabla 6).

Fecha	Data Base	Estrategia de búsqueda	Resultados
20130721	EBSCO	TI (((Rubric* OR ((Scoring OR Rating*) AND Guide) OR (Assessment instrument)))) AND todos los campos ("higher education" OR Universit*) Limitadores - Publicaciones académicas (arbitradas); Fecha en que se publicó desde: 2000-2013; Tipo de publicación: Academic Modos de búsqueda - Booleano/Frase	56
20130719	SCIENCE DIRECT	pub-date > 1999 and TITLE(((Rubric* OR ((Scoring OR Rating*) AND Guide) OR (Assessment instrument)))) and TITLE-ABSTRACT KEY (("higher education" OR Universit*))	12
20130718	Web Of Science	TI=((Rubric* OR ((Scoring OR Rating*) AND Guide) OR (Assessment instrument))) AND TS=("higher education" OR Universit*) Databases=SCI-EXPANDED, SSCI Timespan=2000-2013	63
20140805	Web Of Science Core collection	TI=(((Rubric* OR ((Scoring OR Rating*) AND Guide) OR (Assessment instrument)))) Timespan=2000-2013 Databases=SCI-EXPANDED, SSCI	2.031
20140414	SCOPUS	TITLE(((Rubric* OR ((Scoring OR Rating*) AND Guide) OR (Assessment instrument)))) PUBYEAR AFT 1999 PUBYEAR BEF 2014	1.912
20140414	SCIENCE DIRECT	Title= (((Rubric* OR ((Scoring OR Rating*) AND Guide) OR (Assessment instrument)))) Timespan=2000-2013	387
20140805	EBSCO Business Source Premier	Title (((Rubric* OR ((Scoring OR Rating*) AND Guide) OR (Assessment instrument)))) Timespan=2000-2013 Limitadores - Publicaciones académicas (arbitradas); Modos de búsqueda - Booleano/Frase	75

Tabla 6. Resultados de la estrategia de búsqueda

Tras eliminar los duplicados y seleccionar los artículos aplicando los criterios de inclusión y exclusión con la información de título, palabras clave y resumen, queda un conjunto de 252 referencias útiles (Figura 1). De ellos, como mínimo hay 47 trabajos que se centran en asignaturas o titulaciones universitarias relacionadas con la gestión o la ingeniería. También hay 156 artículos que no ofrecen información sobre la materia donde se han aplicado las rúbricas.

Partiendo de los 252 resultados filtrados, 11 de estas referencias están incluidas en las tres revisiones de literatura localizadas sobre rúbricas (Tabla 7). 5 de los artículos coinciden con Reddy y Andrade (2010), 5 con Jonsson y Svingby (2007) y 7 con Panadero y Jonsson (2013). Algunas de estas referencias están incluidas en más de una de las revisiones (ver Anexo 1). Por otra parte, hemos localizado 241 nuevos artículos relacionados con el uso de rúbricas en docencia universitaria. Por lo tanto, existe material disponible no utilizado en las revisiones sistemáticas sobre rúbricas publicadas hasta la fecha.

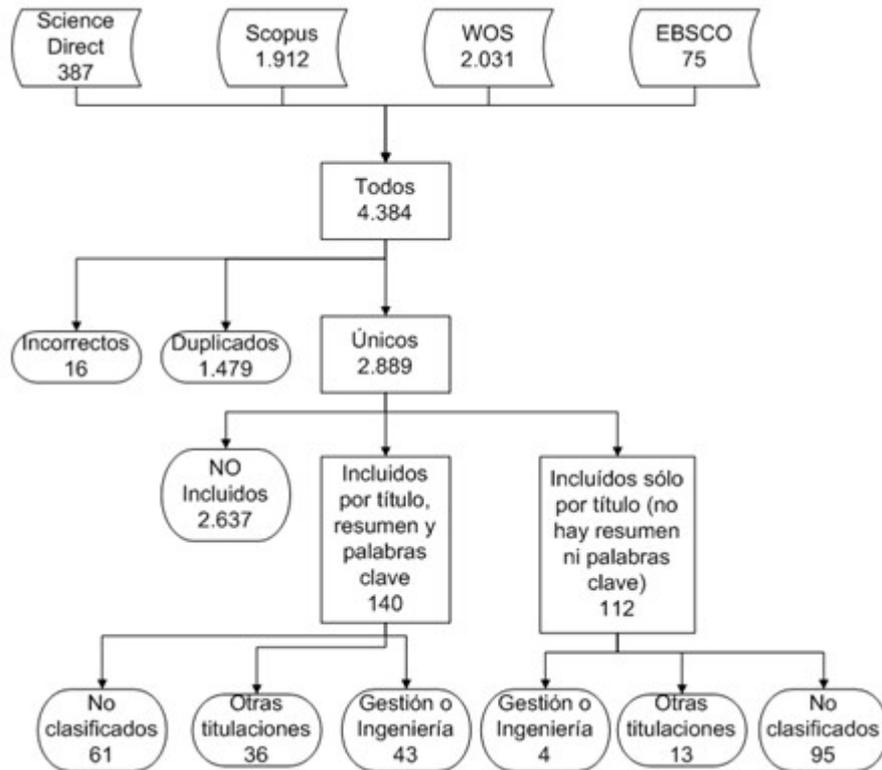


Figura 1. Diagrama QUOROM

	Nuestra búsqueda	(Reddy & Andrade, 2010)	(Jonsson & Svingby, 2007)	(Panadero & Jonsson, 2013)
Referencias incluidas	252	20	75	17
Referencias comunes con las 252 de nuestra búsqueda		5	5	7

Tabla 7. Comparación de resultados de revisiones

3. Conclusiones

En este artículo planteamos 6 preguntas de investigación sobre el uso de rúbricas en las enseñanzas universitarias. Puesto que hay tres revisiones de literatura recientes, hemos comprobado qué respuestas han dado a las 4 primeras preguntas. En general, no existe una respuesta científica concluyente para cómo crear y validar una rúbrica para evaluación de asignaturas universitarias, ni para qué sirven las rúbricas, ni si vale la pena el esfuerzo de desarrollarlas. Existen ciertas recomendaciones para el uso adecuado de rúbricas pero están basadas en evidencias anecdóticas sin una metodología robusta y contrastada para validar esas propuestas. Tampoco es fácil localizar qué asignaturas, ni qué objetos, ni en qué

contextos han sido objeto de investigación las rúbricas, ni qué rúbricas validadas hay disponibles para ellos. En general a pesar del esfuerzo investigador sobre rúbricas, no estamos en condiciones de asegurar prácticamente nada. Existen bastantes propuestas avaladas por el sentido común y ciertas (escasas) investigaciones. La generalización de estas propuestas pasa por una replicación a gran escala de las investigaciones disponibles. Dicha replicación, parece que en estos momentos aún no se ha producido. Precisamente la respuesta a la quinta de nuestras preguntas de investigación ofrece un abanico de posibles líneas de investigación necesarias para hacer avanzar el conocimiento científico sobre el uso de rúbricas. Por último, hemos comprobado que, de las 142 referencias usadas en las revisiones anteriores, pocas informan de las investigaciones desarrolladas en nivel educativo universitario. Además existe material disponible no utilizado anteriormente en revisiones sistemáticas sobre rúbricas. Hemos localizado 241 nuevas referencias sobre rúbricas y es posible que, con ellas, ya se pueda realizar un meta-análisis para dar una respuesta más solvente a alguna de las preguntas de investigación planteadas. A estos resultados habría que añadir material no publicado en revistas indexadas en las bases de datos utilizadas. Por lo que sería recomendable hacer una búsqueda en google scholar para hacer aflorar nuevos falsos negativos (Jarde, Losilla & Vives, 2012).

Por otra parte, en las tres revisiones localizadas, apenas hay evidencias centradas en ingeniería/gestión. Es posible que los resultados derivados de otras disciplinas universitarias o las conclusiones extraídas con alumnos no universitarios puedan extrapolarse a la docencia universitaria de ingeniería o gestión. Sin embargo, nada garantiza que sea así, es una posibilidad que hay que contrastar. Entre los artículos localizados en nuestra búsqueda hay un número importante centrado en ingeniería o gestión, lo que permitiría en el futuro hacer una integración de sus resultados para extraer conclusiones específicas para estas áreas.

Las contribuciones de nuestro trabajo son de interés tanto para los investigadores en docencia universitaria como para los profesores que imparten docencia en universidades y están interesados en la posibilidad de usar rúbricas como herramienta de evaluación. Para ambos, integramos los resultados de tres revisiones de literatura recientes, haciendo un resumen crítico de sus aportaciones y añadiendo matices metodológicos que deberían ser usados en el futuro en la investigación sobre rúbricas para evaluación de asignaturas universitarias. Además de ofrecer guías para la investigación futura, integramos también, la lista de referencias utilizadas en las tres revisiones, suprimiendo duplicados, y añadiendo un conjunto de 241 nuevos artículos localizados. De este modo, presentamos una lista de referencias filtradas con criterios explícitos que pueden ser usadas por los investigadores o profesionales docentes para consulta o para investigación, sin tener que invertir tiempo en el proceso de búsqueda y filtrado de referencias partiendo desde cero.

Este trabajo representa un punto de partida que abre nuevas líneas de investigación. Por un lado es necesario completar la revisión sistemática incorporando material no publicado en revistas indexadas -"grey literature"- . Por otra parte, sería recomendable clasificar las referencias disponibles para comprobar si hay información suficiente para transformar las revisiones narrativas actuales en una serie de meta-análisis que nos permitan integrar las evidencias de una manera más robusta y fiable. En caso de haber material suficiente habría que realizar los meta-análisis y, en caso contrario, tendríamos un mapa de dónde están los nichos de investigación y una estimación de la magnitud de las carencias de estudios concretos. Todo ello permitiría orientar la investigación futura hacia la obtención de estudios suficientes para poder completar los meta-análisis necesarios en el futuro. Por último, es necesario recoger todas las rúbricas validadas existentes y clasificarlas (materia, producto/tarea, objetivo de aprendizaje...) para facilitar el trabajo de creación/adaptación de nuevas rúbricas y evitar que cada profesor tenga que partir desde cero si ya hay algo parecido creado y validado por sus colegas.

Agradecimientos

Este trabajo ha sido realizado con la financiación del programa Erasmus+ de la Unión Europea ["FINCODA" proyecto 554493-EPP-1-2014-1-FI-EPPKA2-KA].

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(1).
<http://dx.doi.org/10.4995/wpom.v5i1.1758>
- BORENSTEIN, M.; HEDGES, L.V.; HIGGINS, J.P.T.; ROTHSTEIN, H.R. (2013). *Introduction to Meta-Analysis*. John Wiley & Sons.
- DELGADO RODRÍGUEZ, M.; LLORCA, J.; DOMÉNECH MASSONS, J.M. (2012). *Estudios para pruebas diagnósticas y factores pronósticos*. Barcelona: Signo.
- DOVAL DIÉGUEZ, E.; VILADRICH SEGUÉS, M.C. (2011). *Desarrollo y adaptación de cuestionarios en el ámbito de la salud*. Laboratori d'Estadística Aplicada i de Modelització (UAB).
- HAIR, J.F.; ANDERSON, R.E.; TATHAM, R.L.; BLACK, W.C. (1999). *Análisis de datos multivariante* (4th ed.). Madrid: Prentice Hall.
- JARDE, A.; LOSILLA, J.M.; VIVES, J. (2012). Suitability of three different tools for the assessment of methodological quality in ex post facto studies. *International Journal of Clinical Health & Psychology*, 12(1): 97-108.
- JONSSON, A.; SVINGBY, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2): 130-144.
<http://dx.doi.org/10.1016/j.edurev.2007.05.002>
- MARÍN-GARCÍA, 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.
- MARÍN-GARCÍA, J.A. (2013). What do we know about the relationship between High Involvement Work Practices and Performance?. *Working Papers on Operations Management*, 4(2): 01-15.
- MARÍN-GARCÍA, J.A.; MIRALLES INSA, C.; MARÍN GARCÍA, P. (2008). Oral Presentation and Assessment Skills in Engineering Education. *International Journal of Engineering Education*, 24(5): 926-935.
- MCGOLDRICK, K.; PETERSON, B. (2013). Using rubrics in economics. *International Review of Economics Education*, 12: 33-47. <http://dx.doi.org/10.1016/j.iree.2013.04.009>
- PANADERO, E.; JONSSON, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9(0): 129-144.
<http://dx.doi.org/10.1016/j.edurev.2013.01.002>

REDDY, Y.M.; ANDRADE, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4): 435-448. <http://dx.doi.org/10.1080/02602930902862859>

VILADRICH SEGUÉS, M.C.; DOVAL DIÉGUEZ, E. (2011). *Medición: fiabilidad y validez*. Laboratori d'Estadística Aplicada i de Modelització (UAB).

WATTS, F.; MARÍN-GARCÍA, J.A.; GARCÍA-CARBONELL, A.; AZNAR-MAS, L.E. (2012). Validation of a rubric to assess innovation competence. *Working Papers on Operations Management*, 3(1): 61-70.

Anexo 1. Comparación de las referencias publicadas a partir del año 2000 e incluidas en las tres revisiones recientes

Cita	(Reddy & Andrade, 2010)	(Jonsson & Svingby, 2007)	(Panadero & Jonsson, 2013)	Nuestra revisión
Andrade, H. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H.J. Andrade & G. J. Cizek (Eds.), <i>Handbook of formative assessment</i> . New York: Routledge.			x	
Andrade, H. 2000. Using rubrics to promote thinking and learning. <i>Educational Leadership</i> , 57, no. 5: 13–18.	x			x
Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. <i>Practical Assessment, Research & Evaluation</i> , 10(3), 1–11. Retrieved from http://pareonline.net/getvn.asp?v=10&n=3 .	x		x	
Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. <i>Theory into Practice</i> , 48(1), 12–19.			x	
Andrade, H., Buff, C., Terry, J., Erano, M., & Paolino, S. (2009). Assessment-driven improvements in middle school students' writing. <i>Middle School Journal</i> , 40(4), 4–12.			x	
Andrade, H., Du, Y., & Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. <i>Assessment in Education: Principles, Policy & Practice</i> , 17(2), 199–214. http://dx.doi.org/10.1080/09695941003696172 .			x	
Andrade, H., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. <i>Educational Measurement: Issues and Practices</i> , 27(2).	x		x	
Andrade, H., Wang, X. L., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. <i>Journal of Educational Research</i> , 102(4), 287–301.			x	
Andrade, H.G. 2001. The effects of instructional rubrics on learning to write. <i>Current Issues in Education</i> 4, no. 4. http://cie.ed.asu.edu/volume4/number4 (accessed January 4, 2007).	x			
Andrade, H.G. 2005. Teaching with rubrics: The good, the bad, and the ugly. <i>College Teaching</i> 53, no. 1: 27–30.	x			
Arter, J., & McTighe, J. (2001). <i>Scoring rubrics in the classroom</i> . Thousand Oaks: Corwin Press Inc..	x	x		
Arter, J., and J. Chappuis. 2007. <i>Creating and recognizing quality rubrics</i> . Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.	x			
Balan, A. (2012). <i>Assessment for learning: A case study in mathematics education</i> . Doctoral dissertation. Malmö University, Malmö, Sweden.			x	
Beeth, M.E., Cross, L., Pearl, C., Pirro, J., Yagnesak, K., & Kennedy, J. (2001). A continuum for assessing science process knowledge in grades K-6. <i>Electronic Journal of Science Education</i> , 5.		x		

Cita	(Reddy & Andrade, 2010)	(Jonsson & Svingby, 2007)	(Panadero & Jonsson, 2013)	Nuestra revisión
Bissell, A. N., & Lemons, P.R. (2006). A new method for assessing critical thinking in the classroom. <i>BioScience</i> , 56, 66–72.		x		
Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. <i>Educational Assessment Evaluation and Accountability</i> , 21, 5–31.			x	
Bolton, C.F. 2006. Rubrics and adult learners: Andragogy and assessment. <i>Assessment Update</i> 18, no. 3: 5–6.	x			
Borko, H., & Stecher, B. (2006). Using classroom artifacts to measure instructional practice in middle school science: A two-statefield test. CSE Technical Report 690. Los Angeles.		x		
Boston, M., & Wolf, M.K. (2006). Assessing academic rigor in mathematics instruction: The development of the instructional quality assessment toolkit. CSE Technical Report 672. Los Angeles.		x		
Boulet, J.R, T.A. Rebbecchi, E.C. Denton, D. Mckinley, and G.P. Whelan. 2004. Assessing the written communication skills of medical school graduates. <i>Advances in Health Sciences Education</i> 9: 47–60.	x			
Brookhart, S.M. (2005). The quality of local district assessments used in Nebraska’s school-based teacher-led assessment and reporting system (STARS). <i>Educational Measurement: Issues and Practice</i> , 24, 14–21.		x		
Brown, G. T. L., Glasswell, K., & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. <i>Assessing Writing</i> , 9(2), 105–121. http://dx.doi.org/10.1016/j.asw.2004.07.001 .		x	x	
Campbell, A. 2005. Application of ICT and rubrics to the assessment process where professional judgement is involved: the features of an e-marking tool. <i>Assessment & Evaluation in Higher Education</i> 30, no. 5: 529–37.	x			
Chi, E. (2001). Comparing holistic and analytic scoring for performance assessment with many-facetraschmodel. <i>Journal of Applied Measurement</i> , 2, 379–388.		x		
Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaf folded peer assessment of writing from instructor and student perspectives. <i>Journal ofEducational Psychology</i> , 98, 891–901.		x		
Choinski, E.; Mark, A. E.; Murphey, M. (2003). Assessment with Rubrics: Efficient and Objective Means of Assessing Student Outcomes in an Information Resources Class. portal: Libraries & the Academy, Vol. 3, nº. 4, pp. 563-575.		x		x
Clare, L., Valdes, R., Pascal, J., & Steinberg, J. R. (2001). Teachers’ assignments as indicators of instructional quality in elementary schools. CSE TechnicalReport545. Los Angeles.		x		
Cohen, E.G., R.A. Lotan, P.L. Abram, B.A. Scarloss, and S.E. Schultz. 2002. Can groups learn? <i>Teachers College Record</i> 104, no. 6: 1045–68.	x			
Denner, P. R., Salzman, S. A., &Harris, L. B. (2002). Teacher work sample assessment: An accountability method that moves beyond teacher testing to the impact of teacher performance on student learning. In Paper presented at the annual meeting of the American Association of Colleges for Teacher Education.		x		
Dochy, F. (2006). A guide for writing scholarly articles or reviews for the Educational Research Review. <i>Educational Research Review</i> , 4, 1–2. Retrieved from http://www.journals.elsevier.com/educational-research-review/ .			x	
Duke, B.L. (2003). The influence of using cognitive strategy instruction through writing rubrics on high school students’ writing self-efficacy, achievement goal orientation, perceptions of classroom goal structures, self-regulation, and writing achievement. Unpublished doctoral dissertation. University of Oklahoma, USA.		x	x	
Dunbar, N.E., Brooks, C.F., & Kubicka-Miller, T. (2006). Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. <i>Innovative Higher Education</i> , 31, 115–128.	x	x		x
Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. <i>Educational Psychologist</i> , 46(1), 6–25.			x	

Cita	(Reddy & Andrade, 2010)	(Jonsson & Svingby, 2007)	(Panadero & Jonsson, 2013)	Nuestra revisión
Flowers, C. (2006). Confirmatory factor analysis of scores on the clinical experience rubric. <i>Educational and Psychological Measurement</i> , 66, 478–488.		x		
Flowers, C.P., & Hancock, D.R. (2003). An interview protocol and scoring rubric for evaluating teacher performance. <i>Assessment in Education: Principles, Policy and Practice</i> , 10, 161–168.		x		
Goodrich Andrade, H., & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. <i>Journal of Educational Research</i> , 97(1), 21–34.			x	
Goodrich Andrade, H. (2001). The effects of instructional rubrics on learning to write. <i>Current Issues in Education</i> , 4(4). Retrieved from http://cie.ed.asu.edu/volume4/number4/ .			x	
Green, R.; Bowser, M. (2006). Observations from the Field: Sharing a Literature Review Rubric. <i>Journal of Library Administration</i> , Vol. 45, nº. 1/2, pp. 185-202.	x	x	x	x
Hafner, J. C., & Hafner, P. M. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. <i>International journal of science education</i> , 25, 1509–1528.	x	x	x	x
Halonen, J. S., Bosack, T., Clay, S., McCarthy, M., Dunn, D. S., Hill Iv, G. W., & Whitlock, K. (2003). A rubric for learning, teaching, and assessing scientific inquiry in psychology. <i>Teaching of Psychology</i> , 30(3), 196–208. http://dx.doi.org/10.1207/s15328023top3003_01 .			x	x
Harrison, J. A., McAfee, H., & Caldwell, A. (2002). Examining, developing, and validating the interview for admission into the teacher education program. In Paper Presented at the Annual Meeting of the Southeastern Region Association for Teacher Educators.		x		
Hickey, D. T., DeCuir, J., Hand, B., Kyser, B., Laprocina, S., & Mordica, J. (2002). Technology-supported formative and summative assessment of collaborative scientific inquiry. <i>Learning & Performance Support Laboratory</i> , University of Georgia.		x		
Johnson, R. L., Fisher, S., Willeke, M. J., & McDaniel, F. (2003). Portfolio assessment in a collaborative program evaluation: The reliability and validity of a family literacy portfolio. <i>Evaluation and Program Planning</i> , 26, 367–377.		x		
Johnson, R. L., Penny, J., & Gordon, B. (2000). The relation between score resolution methods and inter rater reliability: An empirical study of an analytic scoring rubric. <i>Applied Measurement in Education</i> , 13, 121–138.		x		
Johnson, R.L., Penny, J., & Gordon, B. (2001). Score resolution and the inter rater reliability of holistics cores in rating essays. <i>Written Communication</i> , 18, 229–249.		x		
Jonsson, A. (2008). <i>Educative assessment for/of teacher competency</i> . Doctoral dissertation. Malmö University, Malmö, Sweden.			x	
Jonsson, A. (2013). <i>Facilitating productive use of feedback in higher education</i> . <i>Active Learning in Higher Education</i> , 14.			x	x
Jonsson, A.; Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. <i>Educational Research Review</i> , Vol. 2, nº. 2, pp. 130-144.	x	x	x	x
Jonsson, A. (2010). The use of transparency in the “Interactive Examination” for student teachers. <i>Assessment in Education: Principles, Policy & Practice</i> , 17(2), 183–197.			x	
Keiser, J.C., Lawrenz, F., & Appleton, J.J. (2004). Technical education curriculum assessment. <i>Journal of Vocational Education Research</i> , 29, 181–194.		x		
Knight, L. A. (2006). Using rubrics to assess information literacy. <i>Reference Services Review</i> , Vol. 34, nº. 1, pp. 43-55.	x			x
Koul, R., Clariana, R.B., & Salehi, R. (2005). Comparing several human and computer-based methods for scoring concept maps and essays. <i>Journal of Educational Computing Research</i> , 32, 227–239.		x		
Kuhl, J. (2000). A functional-design approach to motivation and self-regulation. In M. Boekaerts, P.R. Pintrich, & M. Zeidner (Eds.), <i>Handbook of self-regulation</i> (pp. 111–169). San Diego, CA: Academic Press.			x	

Cita	(Reddy & Andrade, 2010)	(Jonsson & Svingby, 2007)	(Panadero & Jonsson, 2013)	Nuestra revisión
Lapsley, R., and R. Moody. 2007. Teaching tip: Structuring a rubric for online course discussions to assess both traditional and non-traditional students. <i>Journal of American Academy of Business</i> 12, no. 1: 167–72.	x			
Laveault, D., & Miles, C. (2002). The study of individual differences in the utility and validity of rubrics in the learning of writing ability. In Paper Presented at the Annual Meeting of the American Educational Research Association.		x		
Lindblom-Ylänne, S., Pihlajamäki, H., & Kotkas, T. (2006). Self-, peer- and teacher-assessment of student essays. <i>Active Learning in Higher Education</i> , 7, 51–62.		x		
Lunsford, B. E. (2002). Inquiry and inscription as keys to authentic science instruction and assessment for preservice secondary science teachers. Unpublished doctoral dissertation. USA: University of Tennessee.		x		
Marzano, R. J. (2002). A comparison of selected methods of scoring classroom assessments. <i>Applied Measurement in Education</i> , 15, 249–267.		x		
Matsumura, L. C., Slater, S. C., Wolf, M. K., Crosson, A., Levison, A., Peterson, M., Resnick, L., & Junker, B. (2006). Using the instructional quality assessment tool to investigate the quality of reading comprehension assignments and student work. CSE Technical Report 669. Los Angeles.		x		
McCormick, M. J., Dooley, K. E., Lindner, J. R., & Cummins, R. L. (2007). Perceived growth versus actual growth in executive leadership competencies: An application of the stair-step behaviorally anchored evaluation approach. <i>Journal of Agricultural Education</i> , 48(2), 23–35.			x	
McMartin, F., McKenna, A., & Youssefi, K. (2000). Scenario assignments as assessment tools for undergraduate engineering education. <i>IEEE Transactions on Education</i> , 43, 111–120.		x		
Moni, R.W., E. Beswick, and K.B. Moni. 2005. Using student feedback to construct an assessment rubric for a concept map in physiology. <i>Advances in Physiology Education</i> 29: 197–203.	x			
Moskal, B.M., & Leydens, J.A. (2000). Scoring rubric development: Validity and reliability. <i>Practical Assessment, Research & Evaluation</i> , 7, 71–81.	x	x	x	
Mott, M.S., Etsler, C., & Drumgold, D. (2003). Applying an analytic writing rubric to children’s hypermedia “narratives”. <i>Early Childhood Research & Practice: An Internet Journal on the Development, Care, and Education of Young Children</i> , 5.		x		
Mullen, Y.K. (2003). Student improvement in middle school science. Master’s thesis, University of Wisconsin.		x	x	
Oakleaf, M.J. 2006. Assessing information literacy skills: A rubric approach. PhD diss., University of North Carolina, Chapel Hill. UMI No. 3207346.	x			
Olson, L., Schieve, A. D., Ruit, K.G., & Vari, R.C. (2003). Measuring inter-rater reliability of the sequenced performance inventory and reflective assessment of learning (SPIRAL). <i>Academic Medicine</i> , 78, 844–850.		x		
Osana, H.P., & Seymour, J.R. (2004). Critical thinking in preservice teachers: A rubric for evaluating argumentation and statistical reasoning. <i>Educational Research and Evaluation</i> , 10, 473–498.	x	x		
Pajares, F. (2008). Motivational role of self-efficacy beliefs in self-regulated learning. In D. H. Schunk & B.J. Zimmerman (Eds.), <i>Motivation and self-regulated learning. Theory, research and applications</i> (pp. 111–168). New York: Lawrence Erlbaum Associates.			x	
Panadero, E. (2011). Instructional help for self-assessment and self-regulation: Evaluation of the efficacy of self-assessment scripts vs. rubrics. Doctoral dissertation. Universidad Autónoma de Madrid, Madrid, Spain.			x	
Panadero, E., Alonso-Tapia, J., & Huertas, J.A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. <i>Learning and Individual Differences</i> , 22(6), 806–813. http://dx.doi.org/10.1016/j.lindif.2012.04.007 .			x	

Cita	(Reddy & Andrade, 2010)	(Jonsson & Svingby, 2007)	(Panadero & Jonsson, 2013)	Nuestra revisión
Parkes, K.A. 2006. The effect of performance rubrics on college level applied studio grading. PhD diss., University of Maimi. UMI No. 3215237.	x			
Payne, D.A. 2003. Applied educational assessment. 2nd ed. Belmont, CA: Wadsworth/ Thomson Learning.	x			
Penny, J., Johnson, R.L., & Gordon, B. (2000a). The effect of rating augmentation on inter-rater reliability: An empirical study of a holistic rubric. <i>Assessing Writing</i> , 7, 143–164.		x		
Penny, J., Johnson, R.L., & Gordon, B. (2000b). Using rating augmentation to expand the scale of an analytic rubric. <i>Journal of Experimental Education</i> , 68, 269–287.		x		
Petkov, D., & Petkova, O. (2006). Development of scoring rubrics for IS projects as an assessment tool. <i>Issues in Informing Science and Information Technology</i> , 3, 499–510.	x	x	x	
Pindiprolu, S. S., Lignugaris/Kraft, B., Rule, S., Peterson, S., & Slocum, T. (2005). Scoring rubrics for assessing students' performance on functional behavior assessment cases. <i>Teacher Education and Special Education</i> , 28, 79–91.		x		
Piscitello, M. E. (2001). Using rubrics for assessment and evaluation in art. Master's thesis. Saint Xavier University, Chicago (USA).	x	x	x	
Popp, S. E. O., Ryan, J. M., Thompson, M. S., & Behrens, J. T. (2003). Operationalizing the rubric: The effect of benchmark selection on the assessed quality of writing. In Paper Presented at Annual Meeting of the American Educational Research Association.		x		
Powell, T.A. 2001. Improving assessment and evaluation methods in film and television production courses. PhD diss., Capella University. UMI No. 3034481. Practical Assessment, Research and Evaluation 7, no. 18. Available http://pareonline.net/getvn.asp?v=7&n=18 (accessed December 7, 2006).	x			
Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: All that effort, but what is the effect? <i>Assessment & Evaluation in Higher Education</i> , 35, 277–289.			x	
Quinlan, A.M. 2006. A complete guide to rubrics, assessments made easy for teachers, K-College. USA: Rowman & Littlefield Education.	x			
Ramos, K. D., Schafer, S., & Tracz, S. M. (2003). Validation of the fresno test of competence in evidence based medicine. <i>British Medical Journal</i> , 326, 319–321.		x		
Reddy, Y. M. (2007). Effects of rubrics on enhancement of student learning. <i>Educate</i> , 7(1), 3–17.			x	
Reddy, Y. M.; Andrade, H. (2010). A review of rubric use in higher education. <i>Assessment & Evaluation in Higher Education</i> , Vol. 35, nº. 4, pp. 435-448.			x	x
Reitmeier, C. A., & Vrchota, D. A. (2009). Self-assessment of oral communication presentations in food science and nutrition. <i>Journal of Food Science Education</i> , 8(4), 88–92.			x	
Reitmeier, C.A., L.K. Svendsen, and D.A. Vrchota. 2004. Improving oral communication skills of students in food science courses. <i>Journal of Food Science Education</i> 3: 15–20.	x			
Reynolds-Keefer, L. (2010). Rubric-referenced assessment in teacher preparation: An opportunity to learn by using. <i>Practical Assessment Research & Evaluation</i> , 15(8). Retrieved from http://pareonline.net/getvn.asp?v=15&n=8 .			x	x
Roblyer, M. D., & Wiencke, W. R. (2003). Design and use of a rubric to assess and encourage interactive qualities in distance courses. <i>American Journal of Distance Education</i> , 17, 77–99.		x		
Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. <i>Educational Assessment</i> , 11, 1–31.		x	x	
Schacter, J., & Thum, Y.M. (2004). Paying for high- and low-quality teaching. <i>Economics of Education Review</i> , 23, 411–430.		x		
Schafer, W. D., Swanson, G., Bene, N., & Newberry, G. (2001). Effects of teacher knowledge of rubrics on student achievement in four content ' areas. <i>Applied Measurement in Education</i> , 14, 151–170.		x	x	

Cita	(Reddy & Andrade, 2010)	(Jonsson & Svingby, 2007)	(Panadero & Jonsson, 2013)	Nuestra revisión
Schamber, J. F., & Mahoney, S. L. (2006). Assessing and improving the quality of group critical thinking exhibited in the final projects of collaborative learning groups. <i>The Journal of General Education</i> , 55(2), 103–137. http://dx.doi.org/10.1353/jge.2006.0025 .		x	x	
Schneider, J.F. 2006. Rubrics for teacher education in community college. <i>The Community College Enterprise</i> 12, no. 1: 39–55.	x			
Shaw, J. (2004). Demystifying the evaluation process for parents: Rubrics for marking student research projects. <i>Teacher Librarian</i> , 32, 16–19.		x		
Simon, M., & Forgette-Giroux, R. (2001). A rubric for scoring postsecondary academic skills. <i>Practical Assessment, Research & Evaluation</i> , 7.	x	x	x	
Song, K.H. 2006. A conceptual model of assessing teaching performance and intellectual development of teacher candidates: A pilot study in the US. <i>Teaching in Higher Education</i> 11, no. 2: 175–90.	x			
Stiggins, R.J. 2001. <i>Student-involved classroom assessment</i> . 3rd ed. Upper Saddle River, NJ: Prentice-Hall.	x			
Stoddart, T., Abrams, R., Gasper, E., & Canaday, D. (2000). Concept maps as assessment in science inquiry learning—A report of methodology. <i>International Journal of Science Education</i> , 22, 1221–1246.		x		
Stoering, J.M., & Lu, L. (2002). Combining the national survey of student engagement with student portfolio assessment. In Paper Presented at Annual Meeting of the Association for Institutional Research.		x		
Tierney, R., and M. Simon. 2004. What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels'. <i>Practical Assessment, Research & Evaluation</i> 9, no. 2.	x			
Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. <i>Assessment in Education: Principles, Policy & Practice</i> , 14, 281–294.			x	
Toth, E. E., Suthers, D. D., & Lesgold, A. M. (2002). "Mapping to know": The effects of representational guidance and reflective assessment on scientific inquiry. <i>Science Education</i> , 86(2), 264–286. http://dx.doi.org/10.1002/sce.10004 .		x	x	
Tuñón, J.; Brydges, B. (2006). A Study on Using Rubrics and Citation Analysis to Measure the Quality of Doctoral Dissertation Reference Lists from Traditional and Non traditional Institutions. <i>Journal of Library Administration</i> , Vol. 45, nº. 3/4, pp. 459-481.	x			
van Dinther, M., Dochy, F., & Segers, M. (2010). Factors affecting students' self-efficacy in higher education. <i>Educational Research Review</i> , 6(2), 95–108. http://dx.doi.org/10.1016/j.edurev.2010.10.003 .			x	
Ward, J.R., & McCotter, S.S. (2004). Reflection as a visible outcome for preservice teachers. <i>Teaching and Teacher Education</i> , 20, 243–257.		x		
William, D. (2011). What is assessment for learning? <i>Studies in Educational Evaluation</i> , 37, 2–14.			x	
Williams, L., & Rink, J. (2003). Teacher competency using observational scoring rubrics. <i>Journal of Teaching in Physical Education</i> , 22, 552–572.		x		
Wolters, C.A. (2003). Regulation of motivation: Evaluating an under emphasized aspect of self-regulated learning. <i>Educational Psychologist</i> , 38(4), 189–205.			x	

Intangible Capital, 2015 (www.intangiblecapital.org)

El artículo está con Reconocimiento-NoComercial 3.0 de Creative Commons. Puede copiarlo, distribuirlo y comunicarlo públicamente siempre que cite a su autor y a Intangible Capital. No lo utilice para fines comerciales. La licencia completa se puede consultar en

<http://creativecommons.org/licenses/by-nc/3.0/es/>