

El diseño de rúbricas: algunos aspectos claves

Creating rubrics: some key aspects

Mireia Alcón Latorre

Universidad de Barcelona

malcon@ub.es

José Luis Menéndez Varela¹

Universidad de Barcelona

menendez@ub.edu

Fecha de recepción del artículo: 20 de septiembre de 2018

Fecha de aceptación: 5 de noviembre de 2018

Resumen

El diseño de rúbricas genera no poca literatura que adopta la forma de recomendaciones y consejos a seguir para llevar a cabo el proceso. Tres son los aspectos claves a considerar antes de adentrarse en sus particularidades: Primero, crear las rúbricas en alineación con los programas y contextos en los que se insertan; segundo, afrontar el proceso desde el trabajo colaborativo; y tercero, contemplar en él la participación de los estudiantes. A partir de estos fundamentos, la creación de los contenidos de una rúbrica presenta toda una serie de retos –como la elaboración de sus criterios de evaluación o el establecimiento de un lenguaje claro y efectivo, entre otros– que solo pueden ser afrontados desde el conocimiento experto de un colectivo comprometido con su práctica profesional.

Palabras clave: rúbricas, diseño de rúbricas, educación superior.

Abstract

There is plenty of literature about rubrics design that takes form of recommendations and advices for their creation. There are three key aspects that should be considered before going into the particularities of the process. First, rubrics should be aligned with programmes and contexts to which they belong, second, the process of designing should be faced collaboratively, and third it should contemplate students' participation. Taking into account these considerations, creating the contents of a rubric presents some challenges –for instance drafting its assessment criteria or using a clear and effective language, among others- that can only be tackled from the expert knowledge of a group committed to their professional practice.

Keywords: rubrics, rubrics design, higher education.

¹ Correspondencia: Departamento de Historia del Arte. Facultad de Geografía e Historia. Universidad de Barcelona. C/Montalegre, 6. 08001. Barcelona, España.

Introducción

Las rúbricas son recursos educativos que reportan múltiples beneficios tanto en su uso para la evaluación del desempeño de los estudiantes como para la evaluación de programas académicos (Alcón Latorre, Menéndez Varela y Arbesú García, 2017). La mayor de sus fortalezas radica en el claro establecimiento de unos criterios de evaluación, para los que se diferencian varios niveles de consecución a través de una serie de descripciones cualitativas². Esta característica de las rúbricas, complementada con un uso formativo de las mismas, otorga transparencia a la evaluación y favorece en quien la usa la comprensión de los objetivos de aprendizaje y/o estándares de calidad establecidos. De esta manera, las rúbricas sistematizan los procesos de evaluación y ayudan en la valoración de constructos complejos, como pueden ser determinados desempeños propios de una evaluación por competencias. Este y muchos otros beneficios –como por ejemplo su papel en la mejora del retorno educativo, su contribución en el desarrollo de competencias metacognitivas o su aportación para el desarrollo de un lenguaje profesional compartido– las convierten en uno de los elementos clave para la implementación de lo que se viene llamando el nuevo paradigma educativo.

A pesar de ello, la integración del recurso en los entornos de educación superior es todavía un fenómeno extraordinario (Sridharan, Leitch y Watty, 2015), aún siendo un tema de creciente interés entre la comunidad científica. Esta escasa implementación de las rúbricas es debida a varios motivos: un uso generalizado de sistemas de evaluación tradicionales, la falta de formación específica que ayude al profesorado en su comprensión y utilización, o la falta de una monitorización externa y real en cuanto a estándares de calidad (Atkinson y Lim, 2013). A estas razones cabe añadirles una cierta reticencia respecto al recurso que opaca su reconocimiento y adopción, y que es fruto de ciertas limitaciones tradicionalmente asociadas al recurso. De entre las más recurridas se encuentran la consideración de que son instrumentos rígidos o restrictivos, las dudas sobre su fiabilidad como instrumento de evaluación o la creencia de que no son útiles para afrontar determinadas competencias.

Muchas de estas limitaciones asociadas al trabajo con rúbricas tienen su origen en un incorrecto planteamiento de su diseño y/o aplicación, ya que estos procesos tienen una estrecha relación con la mayor o menor efectividad del recurso. Por ejemplo, una rúbrica no alineada con los programas en los que se inserta fallará a la hora de conseguir grados aceptables de validez (Wolf y Goodwin, 2007), de ahí que se deba prestar especial atención y cuidado a este y a otros aspectos que analizaremos posteriormente. De hecho, el grueso de los retos que plantea el trabajo con rúbricas se encuentra en sus procesos de diseño y aplicación, de ahí que numerosas considera-

² Esta composición es clave, ya que diferencia una rúbrica de otros sistemas de evaluación. Más allá de sus tres componentes –criterios de evaluación, descripciones cualitativas y niveles de consecución– la particularidad que aporta valor y utilidad a las rúbricas es la inclusión de descripciones para cada nivel de consecución, generando diferentes gradaciones de calidad.

ciones a este respecto llenen buena parte de la literatura relacionada y generen cierto debate entre la comunidad científica.

En este sentido, el artículo pretende ser un compendio de las aportaciones que diversos autores han realizado en torno al diseño de rúbricas de cara a esclarecer ciertos asuntos y dar uniformidad a lo recogido por la literatura especializada. Para ello se abordarán los tres aspectos claves que deberían regir el proceso de diseño de rúbricas, para posteriormente tratar cuestiones más prácticas sobre los pasos o fases que se deben seguir a la hora de enfrentarse a él. Es importante matizar que, a pesar de ser procesos distintos, en ocasiones se habla tanto de diseño como de aplicación de rúbricas por estar estrechamente ligado el uno con el otro: cuando se diseña una rúbrica ya se contempla su aplicación y mientras se aplica se someten a valoración aspectos de su diseño.

Aspectos claves en el diseño de rúbricas

Existe cierto acuerdo entre la comunidad científica sobre cuáles son los aspectos claves a considerar a la hora de afrontar un proceso de diseño de rúbricas. El primero de ellos es la importancia de diseñarlas en alineación con los sistemas de evaluación en los que se insertan y, a su vez, con los programas donde estos se contextualizan. El segundo es la necesidad de afrontar la tarea del diseño desde la colaboración y la alianza con equipos afines y comprometidos. El tercero y vinculado con este último, el fomentar la participación de los estudiantes tanto en el diseño como en la aplicación de las rúbricas. A continuación analizamos cada una de estas consideraciones.

Sobre la alineación de las rúbricas

Un diseño de rúbricas que favorezca la efectividad del recurso debe empezar por contemplar la relación de los contenidos que las componen con los contextos en los que estas se insertan. La recomendación de la literatura es que las rúbricas se diseñen y apliquen pensando en su articulación dentro del engranaje de los programas sobre los que actúan, por lo que deben partir por integrar los objetivos de aprendizaje o estándares de calidad que se hayan establecido. En efecto, la previa definición de la finalidad y objetivos de la práctica educativa es el primer paso en el ejercicio por lograr una alineación constructiva de los programas, que es la que contempla el vínculo entre los procesos de enseñanza, aprendizaje y evaluación (Biggs, 1996). Una vez estos estén definidos cabría establecer las actividades o metodologías orientadas a su cumplimiento para, finalmente, elaborar sólidos sistemas que permitieran la evaluación de los resultados; las rúbricas deberían quedar articuladas entre estos procesos a modo de eje entre ellos (Timmerman, Strickland, Johnson y Payne, 2011; Wilson y Onwuegbuzie, 1999; Wolf, Connelly y Komara, 2008).

El asunto de la alineación de las rúbricas con los programas es percibido, a menudo, como una de las debilidades del trabajo con rúbricas (McKeown, 2011) por su dificultad a la hora de afrontarlo.

Sin embargo, cumplir con el principio de alineación es un aspecto fundamental para que estas puedan desplegar todo su potencial educativo y, de esta manera, puedan aportar todos los beneficios que se les atribuye. Según Moskal y Leydens (2000), una rúbrica correctamente articulada debe incluir en sus criterios todos los objetivos de aprendizaje o estándares de calidad establecidos y, a su vez, ninguno de sus criterios debe quedar desvinculado de los antedichos objetivos o estándares, de lo contrario se compromete la adecuación de la rúbrica. Este trabajo comporta una cuidada planificación del diseño de la rúbrica para integrarlo con los programas, de la que dan testimonio algunos estudios (ej: Allen y Kelly-Riley, 2005; Atkinson y Lim, 2013; Etkina et al., 2006). Es por ello que, con la premisa de la alineación como prioridad, las rúbricas se reconocen como recursos útiles para afrontar el diseño de programas, ya que en su conceptualización –y posterior aplicación– puede detectarse el grado de alineación de los programas al favorecer la identificación de posibles carencias al respecto.

Sobre el trabajo colaborativo

Una mirada atenta y crítica a las rúbricas las reconoce como un recurso que va más allá de la mera evaluación, y percibe cuán importantes son el compromiso y la colaboración en su diseño y aplicación. De hecho, Griffin (2009) las define como proyecciones del compromiso mutuo del colectivo profesional que las crea, como materializaciones de un trabajo colaborativo que busca el consenso de las diferentes representaciones de sus creadores en aras de una mejora de su práctica profesional. Así concebidas, las rúbricas albergan un gran valor en tanto que recogen los conocimientos compartidos de un grupo de expertos en torno a una determinada disciplina o ámbito profesional. La coherencia con estas consideraciones mueve a que la creación de una rúbrica sea, por tanto, fruto de la colaboración, más aún si se tiene en cuenta que el conocimiento común sobre estándares de calidad u objetivos de aprendizaje goza de más valor e impacto que el que pueda tener un único experto a título individual (Reddy, 2011).

A la hora de afrontar este trabajo colaborativo, los procesos de moderación resultan claves para poder llegar a un consenso de los juicios de cada evaluador. Ellos suponen oportunidades donde los participantes pueden discutir, explorar y defender sus posturas en torno a aquello que se evalúa. En este sentido, conseguir un consenso de las diferentes representaciones es esencial para poder emitir juicios significativos y robustos, además de ser esencial para asegurar la fiabilidad en los procesos de evaluación (Belanger et al., 2015; Slomp, 2015). Este proceso tampoco está exento de retos, dada la dificultad de conseguir un entendimiento compartido sobre los diversos criterios y gradaciones de calidad que debe contemplar una rúbrica. (Sridharan et al., 2015). Prueba de ellos son los diversos estudios que contemplan los recursos utilizados para lograr consensos en los procesos de diseño de sus rúbricas (ej: Frazee, 2001; Moni R. y Moni, 2008; Timmerman et al., 2011).

Sobre la participación de los estudiantes

Es mucha la literatura que asevera la idoneidad de la participación de los estudiantes en la aplicación de las rúbricas, especialmente si se trata de la evaluación de su desempeño y mediante procesos de autoevaluación y/o evaluación entre iguales (ej: Belanger et al., 2015; Franklin y Smith, 2015; Frichlin, 2006; Hegler, 2003; Maxwell, 2010; Menéndez Varela, 2013; Montgomery, 2002; Oakleaf, 2008; Shaw, 2014). Los beneficios que puede reportar esta práctica giran en torno al desarrollo de competencias metacognitivas, siempre y cuando esta se lleve a cabo en contextos de evaluación formativa y habiendo trabajado un entendimiento compartido entre docentes y estudiantes sobre los contenidos de la rúbrica. De igual manera es recomendable que los estudiantes participen en la aplicación de rúbricas destinadas a la evaluación de programas, pues no olvidemos que los programas tienen sentido en cuanto se conceptualizan y diseñan pensando en el aprendizaje del estudiante (Alcón Latorre, Menéndez Varela y Arbesú García, 2017), y por tanto sus valoraciones pueden ser valiosas para mejorar ciertos aspectos relativos a la calidad de los programas.

Sin embargo, la participación de los estudiantes en los procesos de diseño de rúbricas es un tema mucho menos explorado, quizás debido a la planificación y esfuerzo extra que esto supone por parte del docente o promotor de la experiencia. Aún así, varios estudios exponen claramente la posibilidad y conveniencia de que esta participación se lleve a cabo (ej: Gezie et al., 2012; Hegler, 2003; Raposo Rivas, Cebrián de la Serna y Martínez-Figueira, 2014; Reddy y Andrade, 2010; Wilson y Onwuegbuzie, 1999), muy especialmente si las rúbricas van a ser destinadas a evaluar su desempeño. Fomentar la implicación del estudiantado en el diseño de rúbricas no está exento de retos. Entre los más importantes se encuentra el de suscitar la motivación suficiente para que generen ideas propias sobre conceptos con los que normalmente no están familiarizados y el de describir productos de calidad cuando su información sobre estos productos no es la de un experto (Lewis, Berghoff y Pheaney, 1999). Las expectativas de una respuesta positiva por parte de los estudiantes no son siempre cumplidas y esto puede frustrar muchas de las iniciativas. Cabe tener en cuenta, sin embargo, que estos se enfrentan a un reto al que seguramente nunca han sido expuestos anteriormente y que, además, puede ser percibido como una prueba de su conocimiento. Una de las posibles soluciones que contempla la literatura para facilitar la comprensión de este ejercicio por parte del estudiante –y que sirve también tanto para afrontar el diseño de la rúbrica como su aplicación, incluso en manos de evaluadores expertos– es la utilización de diferentes muestras que sirvan como ejemplo de las diferentes gradaciones de calidad de los desempeños o estándares de calidad a evaluar. Estas pueden ayudar a los estudiantes a entender cuál es la finalidad de la rúbrica y qué es lo que pueden contemplar sus diferentes descriptores cualitativos.

Los procesos mediante los que se fomenta esta participación parten normalmente de discusiones abiertas en el aula, donde el docente actúa como dinamizador y organizador de las aportaciones de los estudiantes. En otras ocasiones, estas aportaciones se inician de manera individual por

cada estudiante o mediante su distribución en varios grupos de trabajo. Posteriormente, las ideas generadas son puestas en común y sintetizadas por el docente o evaluador al cargo en sesiones colectivas. Se trata entonces de rúbricas negociadas que representan el consenso de la opinión de los estudiantes que, a su vez, se encuentra en consenso con la del docente.

Fases en el diseño de rúbricas

Existe cierto acuerdo en la literatura en cuanto al protocolo a seguir a la hora de afrontar el diseño de una rúbrica (Knight, 2006). A grandes rasgos, este quedaría resumido en tres grandes bloques: conceptualización del recurso, creación –mediante el establecimiento de sus componentes– y testeo –en algunos casos, solo en entornos simulados; en otros, en entornos reales. De entre las recomendaciones recogidas en la literatura sobre el diseño y aplicación de rúbricas, los bloques anteriores pueden desarrollarse en la siguiente secuencia (Menéndez Varela, 2013; Reddy, 2011):

- Definición de la finalidad y propósito del sistema de evaluación. Establecimiento de los objetivos concretos de la rúbrica y alineación con los programas.
- Revisión de la literatura existente y búsqueda de modelos.
- Articulación de los resultados de aprendizaje, en caso de una evaluación del desempeño, o de los estándares de calidad a contemplar por la rúbrica en forma de criterios de evaluación. Desglose de estos en los descriptores cualitativos y niveles de consecución correspondientes.
- Tipificación de la rúbrica. Test de validez y fiabilidad. Aplicación de los ajustes necesarios para su homogeneización.

Los dos primeros puntos compondrían la fase de conceptualización, pues en ella se llevan a cabo las reflexiones e investigaciones iniciales necesarias para fundamentar el diseño de la rúbrica. El tercer punto correspondería a la fase de creación, en ella se establecería el contenido de los elementos principales de la rúbrica y se le proporcionaría el formato que mejor se adecuara a sus propósitos. Por último, el cuarto punto es el que conformaría la fase de testeo y tendría como finalidad probar la adecuación de la rúbrica diseñada para los propósitos establecidos. El primer punto es el que se vincula con lo tratado anteriormente respecto a la alineación de las rúbricas, la recomendación básica es la integración de estos recursos dentro de los procesos de enseñanza, aprendizaje y evaluación. El segundo punto está sujeto a las particularidades de los contenidos a tratar y del contexto educativo. En estas circunstancias, la recomendación principal es que la consulta de modelos de rúbricas similares se acompañe de un análisis cauteloso de en qué medida esos modelos sirven a la finalidad y usos específicos del nuevo contexto. Este análisis deberá determinar si son útiles o no esos modelos; lo habitual es que lo sean solo parcialmente, lo que obliga a introducir toda una serie de adaptaciones. Aunque la utilización de un modelo genérico de rúbrica es rápido y fácil, el diseño de una rúbrica que responda específicamente a las característi-

cas distintivas del objeto de evaluación, actividades programadas y prácticas profesionales concretas ofrecen sin ninguna duda los mejores resultados. Por lo que respecta al tercer y cuarto punto existen numerosas reflexiones de la literatura que adoptan la forma de recomendaciones o consejos a seguir. A continuación se expone una compilación que reúne las directrices básicas aportadas por todas ellas.

Establecimiento de los criterios, descriptores cualitativos y niveles de consecución de una rúbrica

Por lo general, la reflexión sobre cómo abordar esta fase del diseño parte de la evaluación del desempeño y se fundamenta en las pautas establecidas por reconocidos manuales de evaluación, como el de Brookhart (1999), Huba y Freed (2000) o Wiggins (1998). Estas referencias suponen el marco teórico sobre el que se sustentan muchas de las recomendaciones sobre el diseño de rúbricas y la discusión de su contenido. La aportación de Wiggins (1998), basada en una serie de preguntas a plantearse en el diseño de rúbricas para la evaluación del desempeño, es quizás de los recursos más referenciados directa o indirectamente. El autor propone responder a las preguntas de qué criterios evidenciarán el desempeño del estudiante, cómo se valorará la calidad del mismo, cómo se distinguirán y nombrarán los diferentes niveles de calidad y cómo se determinará la puntuación correspondiente de una manera válida y fiable. En efecto, el mayor reto que plantea el diseño de rúbricas es el encontrar indicadores claros y válidos para evidenciar la consecución de los objetivos o estándares establecidos y sobre los que, además, puedan existir grados aceptables de acuerdo y conformidad entre los evaluadores que las utilicen (Dahm, y Newell, 2002; Hammill, Best y Anderson, 2015; Oakleaf, Millet y Kraus, 2011; Newell J.A., Schlitz et al., 2009; Wolf et al., 2008). Tal ejercicio de claridad y transparencia suele ser visto como un verdadero desafío por parte de la comunidad docente, ya que supone explicitar un conocimiento que suele ser tácito (Allen y Kelly-Riley, 2005; Bell, Mladenovic y Price, 2013; Chan, 2015) y puede convertirse en un obstáculo para el desarrollo del proceso³. El esfuerzo considerable y la inversión de tiempo que este ejercicio supone, así como el desconocimiento de herramientas y dinámicas que agilicen estos procesos, es uno de los retos que con mayor frecuencia se le atribuyen al uso de rúbricas.

Para afrontarlo y poder responder a las cuestiones planteadas por Wiggins (1998), un recurso recurrente son los cuatro pasos del diseño de rúbricas propuestos por Stevens y Levi (2005): reflexión, enumeración, clasificación y creación. Con el término reflexión se alude a un análisis en profundidad del proceso de evaluación a afrontar, lo que ayudaría a establecer un listado sobre los aspectos claves a considerar. Una vez concretados estos criterios, la tarea de identificar rasgos u observaciones que los evidencien resultaría más sistemática, de manera que se facilitarí

³ Sobre las posibilidades de articulación de un conocimiento tácito véase Bird y Yucel (2013) y Menéndez-Varela y Groggi-Giralt (2018a).

su detalle en descriptores cualitativos y su clasificación en diferentes niveles de consecución. Finalmente, y con los contenidos de la rúbrica establecidos, ya se podría dar forma final a la rúbrica.

Criterios o dimensiones de evaluación.

La enumeración de los criterios o dimensiones a considerar en la rúbrica comporta un complejo proceso de concreción que además debe llevarse a cabo en alineación con los programas⁴. Para abordar esta dificultosa tarea se suelen organizar sesiones en las que los participantes trabajan de manera colaborativa y se apoyan en la literatura para reforzar las ideas aportadas. De este ejercicio surgen procesos de codificación que permiten establecer categorías sobre las que se construyen el grupo de criterios acordados. Se recomienda partir de un mínimo de tres criterios y llegar hasta un máximo de seis. Esta horquilla parece encontrar el balance entre su exceso, que podría convertir la evaluación en un proceso abrumador, y su defecto, que dificultaría la evaluación al carecer de aspectos claves sobre los que apoyar el proceso (Menéndez Varela, 2013; Menéndez Varela y Gregori Giralt, 2016a; Wolf y Stevens, 2007).

Descripciones cualitativas y niveles de consecución.

La tarea de definir las descripciones cualitativas o los indicadores de calidad correspondientes a cada criterio o dimensión va más allá de la mera enumeración. Se trata de su definición en detalle y su expansión en diferentes gradaciones de calidad, lo que requiere de un profundo conocimiento y un análisis minucioso del objeto de evaluación en cuestión. El reto es definir tanto aquellas características propias de un nivel experto o máximo y las propias de un nivel no competente o mínimo, así como todo el rango de posibilidades susceptible de identificarse entre estos dos extremos de consecución (Allen y Tanner, 2006; Arancibia Aguilera, 2014). Moskal (2000), citando a Brookhart (1999), propone empezar identificando las cualidades necesarias para que el desempeño del estudiante sea considerado experto o competente, estas conformarían el nivel de mayor consecución en la rúbrica. Una vez hecho esto, sugiere continuar por centrar la atención en definir el nivel de menor consecución de la rúbrica, respondiendo a qué tipo de desempeño demostraría un conocimiento muy limitado de los criterios a evaluar. Según la autora, el contraste existente entre el nivel máximo y mínimo facilitaría la definición del nivel medio, con lo que se conformarían tres niveles de consecución. Este ejercicio de comparación entre niveles volvería a ser necesario en el caso de querer conformar más gradaciones de calidad, hasta alcanzar el número de niveles deseados o hasta no poder identificar distinciones relevantes entre ellos.

En cuanto al número adecuado de niveles que deberían conformar una rúbrica parece no haber consenso en la literatura, quizás porque el efecto que este puede ocasionar en la efectividad y uso

⁴ El término *dimensiones* es el más utilizado por la literatura especializada en los últimos años. El propósito de su utilización es diferenciar terminológicamente cada componente de una rúbrica y no confundirlos entre sí. Las descripciones cualitativas que desgranar los criterios de evaluación de una rúbrica no dejan de ser, también, criterios de evaluación solo que de contenido más detallado.

de las rúbricas no es un tema que suscite especial interés (Reddy, 2011). Wolf y Stevens (2007) y Suskie (2004) son de los pocos que establecen recomendaciones específicas a este respecto, sugiriendo, al igual que con los criterios de evaluación, un número entre tres y seis niveles. Algunos autores consideran que resulta complicado para un individuo distinguir más allá de tres niveles de consecución de una manera que extraiga información relevante de esa distinción. Básicamente, esta supone una identificación de una escala manejable en la que el rasgo a evaluar “está presente”, “está presente de alguna manera” o “no está presente” (Allen y Tanner, 2006). Sin embargo, un número impar parece no ser adecuado, ya que supone dejar un nivel en un término medio (Pagano, Bernhardt, Reynolds, Williams y McCurrie, 2008); este nivel intermedio podría adolecer de pocos atributos específicos o facilitar el posible sesgo del evaluador a optar por los valores centrales de la escala de valoración.

La introducción de niveles de consecución en una rúbrica implica añadir sus descripciones cualitativas correspondientes, por tanto, a más niveles más información sobre el fenómeno evaluado se obtendrá. Esto puede reportar mayores beneficios a la hora de facilitar un retorno de información de calidad a los diferentes grupos de interés que puedan estar implicados. Sin embargo, más niveles de consecución implican también un mayor esfuerzo y más dificultades de acotación a la hora de diseñar la rúbrica, mayor tiempo de reflexión a la hora de aplicarla y mayores posibilidades de que la fiabilidad se vea comprometida. Por el contrario, un menor número de niveles facilitaría la tarea del personal evaluador tanto en el diseño y aplicación de la rúbrica y favorecería la posibilidad de obtener mayores grados de fiabilidad, pero comprometería la calidad de la información proporcionada por la rúbrica, aspecto que debería ser cuidadosamente sopesado. En definitiva, el número de niveles de consecución y descriptores cualitativos dependerá de la finalidad que se haya establecido para la rúbrica (Menéndez Varela, 2013; Wolf y Stevens, 2007), de ahí la importancia de este paso previo en el diseño de la misma.

El proceso de calificación: de lo cualitativo a lo cuantitativo.

Otro de los retos específicos que plantea el establecimiento de los distintos niveles de calidad de una rúbrica es su conversión en puntuaciones. Algunos docentes encuentran especialmente difícil valorar mediante calificaciones específicas los indicadores que han observado en el objeto o sujeto evaluado (Sridharan et al., 2015). La dificultad radica en interpretar la información cualitativa que proporciona la rúbrica de una manera cuantitativa, ya que puede suceder que no existan equivalencias exactas entre las diferentes gradaciones de calidad de la rúbrica y las puntuaciones a asignar (Newell J.A. et al., 2002; Petkov y Petkova, 2006). El asunto es una cuestión no resuelta con unanimidad, pues pasa por determinar en cada caso cuál es la solución más conveniente de acuerdo a la finalidad de la evaluación y a los consensos establecidos en el trabajo colaborativo de diseño de la rúbrica. No obstante, este no debería ser un asunto de especial relevancia, de considerar las rúbricas como recursos educativos de carácter formativo. Tal y como apunta Montgomery (2002), si las rúbricas han sido integradas dentro de los procesos de aprendizaje, asignar

una puntuación es solo un pequeño paso más, un acto que sirve de vehículo para informar a los grupos de interés pero que no contiene tanta carga significativa sobre el progreso del aprendizaje de los estudiantes.

El tema del lenguaje.

Por último, otro de los aspectos que sí parece resultar clave debido a su influencia en el entendimiento de la rúbrica es el tema del lenguaje. Encontrar los términos adecuados para definir de manera comprensible los criterios y sus correspondientes descriptores es una tarea de gran complejidad (McNamara y Burton, 2009; Reddy y Andrade, 2010). El grado de claridad y especificidad con la que se muestran los criterios de una rúbrica es un tema que protagoniza muchos de los debates en torno al diseño y aplicación del recurso. Por lo general, se considera que los descriptores cualitativos de una rúbrica deben ser explícitos y no contener ambigüedades que pudieran dar lugar a equívocos; esto es de vital importancia de cara a conseguir un entendimiento compartido del instrumento. Para ello, se recomienda no emplear un lenguaje evasivo, como pudieran ser términos que implican juicios de valor o términos abstractos que aludan a evidencias no medibles (Lundstrom, Diekema, Leary, Haderlie y Wendy, 2015; Moskal, 2000). En efecto, las descripciones cualitativas deberían aludir a cualidades observables y directas, esto es, cualidades fundamentadas empíricamente e identificables en aquello que se pretende evaluar (Jonsson, 2014, Simon y Forgette-Giroux, 2000). A este respecto, la literatura se decanta por la utilización de descriptores ricos y detallados, redactados de una manera consistente y secuencial, de manera que promuevan la reflexión del evaluador, le ayuden a dar forma a lo observado y a emitir juicios fundamentados (Knight, 2006; Lund, 2006; McNamara y Burton, 2009; Stellmack, Konheim-Kalkstein, Manor, Massey y Schmitz, 2009). No obstante, parece no ser muy recomendable que los criterios sean tan específicos que dificulten su identificación en las múltiples casuísticas de un proceso de evaluación (Anglin L., Anglin, Schumann y Kaliski, 2008). Este es uno de los mayores retos en la redacción de las descripciones cualitativas: volver a encontrar ese equilibrio entre la generalidad del lenguaje que, aunque puede incrementar la aplicabilidad de la rúbrica también puede aumentar su ambigüedad, y la especificidad o minuciosidad, que puede provocar que la rúbrica sea más difícil de manejar pero puede, a su vez, incrementar su grado de fiabilidad (Reddy, 2011). El trabajo colaborativo, la participación de los estudiantes, la publicación de la rúbrica con antelación para trabajar su entendimiento y el testeo de su validez y fiabilidad son aspectos que contribuyen a refinar y ajustar el lenguaje de la rúbrica y así superar con éxito esta difícil tarea de diseño.

Tipificación de las rúbricas

El proceso de diseño de una rúbrica no finaliza con el establecimiento de sus componentes, sino que se extiende a probar su manejabilidad en manos de los evaluadores que la testan; así como a probar su efectividad para los propósitos para los que ha sido creada. Sólo de este modo se apuntala su transferencia a otros usuarios y contextos. Bien es cierto que las mismas dinámicas de

trabajo que comporta el propio desarrollo de la rúbrica ya implican procesos de tipificación de su uso –especialmente mediante las sesiones de moderación–, así como de alineación con los objetivos establecidos. Sin embargo, comprobar el funcionamiento del recurso en fases piloto, y simulando lo máximo posible los contextos educativos en los que será aplicado, resulta indispensable para avanzar su implementación en entornos reales e identificar las posibles carencias que este pueda tener. De cualquier modo, el diseño de una rúbrica, especialmente si esta se prevé potenciar como herramienta formativa, nunca puede considerarse completo del todo. Las rúbricas deben estar en continuo desarrollo, ya que el contexto en el que se integran las suele someter a una constante actualización de sus contenidos, muy especialmente si se trata de rúbricas para la evaluación del desempeño de los estudiantes (Allen y Tanner, 2006; Wilson y Onwuegbuzie, 1999).

Para probar el grado de manejabilidad y efectividad de la rúbrica, y al margen de los consensos iniciales a los que se haya llegado para diseñarla, es imprescindible asegurarse de que existe un entendimiento compartido de la misma por parte de los evaluadores que la validan (Reddy, 2011), normalmente un grupo más extenso al de los promotores de su diseño. Para ello, se suelen llevar a cabo sesiones formativas en las que se clarifican los contenidos de las rúbricas a utilizar y se tipifica su aplicación. Estas sesiones combinan, por lo general, explicaciones teóricas sobre la naturaleza de la rúbrica con ejercicios prácticos en los que se aplica el recurso, normalmente sobre una muestra seleccionada similar o igual a aquello que se pretende evaluar y representativa de diferentes niveles de consecución. Posteriormente, de los resultados conseguidos se extraen inferencias que son sometidas a discusión con el propósito final de valorar el proceso de aplicación y ajustar o clarificar los aspectos necesarios.

El hecho de que sea conveniente una formación en rúbricas para facilitar su aplicación da cuenta de que hablamos de recursos complejos, que requieren de tiempo y trabajo para aprender a utilizarlos. Esta condición, cuya revelación para algunos evaluadores y docentes suele suceder en las sesiones formativas, es otra de las causas que más críticas genera (Adedoyin, 2013; Bekki, Bernstein, Fabert, Gildar y Way, 2014; Thaler, Kazemi y Huscher, 2009), y que a menudo desmotiva a los usuarios potenciales. No obstante, convendría preguntarse cuál es la alternativa en la evaluación de fenómenos complejos como lo son, por ejemplo, los aprendizajes de los estudiantes –alternativa que no puede consistir en una práctica evaluadora arbitraria–, qué tiempo y esfuerzo supone el diseño de ese sistema alternativo, y cuál es la validez y fiabilidad de dicho sistema de evaluación. Por otro lado, no hay que olvidar que las sesiones formativas sobre la aplicación de las rúbricas no solo tienen como propósito reducir las discrepancias que puedan surgir entre los evaluadores y, por tanto, aumentar el grado de consistencia de las puntuaciones resultantes de la evaluación (Eshun y Osei-Poku, 2013; Reddy y Andrade, 2010; Stellmack et al., 2009). Estas sesiones, en la medida en que familiarizan a los participantes tanto con la rúbrica como con la complejidad del objeto de evaluación, incrementan el conocimiento y la pericia del evaluador, impulsan la reflexión sobre las prácticas profesionales y promueven una mejora compartida de los

procedimientos y recursos que les son propios. Este aspecto adquiere más relevancia todavía si la rúbrica va a ser utilizada por los estudiantes en procesos de autoevaluación y evaluación entre iguales. Tipificar la aplicación de la rúbrica, por tanto, es un paso imprescindible para garantizar la fiabilidad de la rúbrica y, de esta manera, contribuir a la construcción de un recurso educativo y evaluador robusto y válido.

Sobre validez y fiabilidad en el trabajo con rúbricas

El proceso de tipificación de una rúbrica puede considerarse el último paso en su diseño y el que completa la elaboración del recurso. Sin embargo, la conclusión de este proceso no implica automáticamente que el sistema de evaluación en el que se inserta la rúbrica sirva para emitir inferencias válidas para los propósitos que fue creado. La validez y la fiabilidad son dos de los principios básicos que deben contemplarse, si se pretende llevar a cabo una evaluación rigurosa y de calidad (Docktor et al., 2016; Rezaei y Lovorn, 2010). Resulta del todo conveniente validar los sistemas de evaluación basados en rúbricas considerando ambas propiedades en el análisis de los resultados obtenidos de su aplicación. De no arrojar resultados positivos, el sistema de evaluación debería ser revisado y rediseñado.

La validez es un juicio sobre la idoneidad de las inferencias extraídas de los resultados de evaluación, esto es, el grado con el que el sistema de evaluación cumple con la finalidad para la que fue creado (Moskal y Leydens, 2000). En el caso de una evaluación del desempeño, un sistema de evaluación es válido cuando promueve los aprendizajes que mide. Este aspecto es de sumo interés y hay que entenderlo en su literalidad. En efecto, aunque análisis estadísticos habituales en el estudio de la validez puedan arrojar datos que refrendan la adecuación de los resultados obtenidos a los estándares de calidad del objeto de evaluación, no necesariamente prueban que se haya producido un progreso en el aprendizaje (Menéndez-Varela y Gregori-Giralt, 2016b). Aunque la validez en un contexto educativo valora la adecuación de las inferencias extraídas de las puntuaciones de los estudiantes para que diversos agentes involucrados reconozcan la consecución de los resultados de aprendizaje establecidos, es imprescindible que el análisis de la validez evidencie la existencia de un crecimiento cognitivo de los estudiantes (Menéndez-Varela y Gregori-Giralt, 2016b). Así como la consistencia es uno de los principales factores que explican la fiabilidad, en la validez estos factores giran en torno al valor y utilidad educativos de las deducciones obtenidas de la aplicación del sistema de evaluación. Por todo ello, un juicio sobre la validez es fundamentalmente una reflexión sobre la oportunidad de los retos de aprendizaje establecidos y la adecuación del entorno didáctico particular para abordarlos y resolverlos satisfactoriamente.

Por su parte, la fiabilidad es un juicio sobre el uso correcto que se hace de un sistema de evaluación; un uso correcto que se muestra en el grado de concordancia de las puntuaciones otorgadas por los evaluadores. Dicha concordancia manifiesta que existen representaciones compartidas de los evaluadores acerca del sistema de evaluación y una percepción similar de los fenómenos eva-

luados (Menéndez-Varela y Gregori-Giralt, 2018b). No obstante, el juicio de la fiabilidad depende también de la adecuación del sistema de evaluación al desempeño bajo escrutinio, la adecuación del número de pruebas de evaluación y su distribución en el tiempo a los resultados de aprendizaje perseguidos, y las diferentes reacciones de los estudiantes a las actividades, criterios y ocasiones de evaluación (Menéndez-Varela y Gregori-Giralt, 2018b). La fiabilidad depende también de estos factores porque también ellos afectan al grado de concordancia obtenido entre los evaluadores. Por ejemplo, un sistema de evaluación que no valora lo que se pretende no solo incurre en un déficit de validez sino que también penaliza la fiabilidad, en la medida en que podría generar perplejidad, confusión y arbitrariedad entre los evaluadores. De la interacción de estos cuatro factores depende que las puntuaciones representen diferencias verdaderas en el desempeño de los estudiantes, independientemente de cuándo y en qué circunstancias se realizaron las pruebas, y de quiénes, cuándo y en qué circunstancias puntuaron las respuestas del estudiante (Menéndez-Varela y Gregori-Giralt, 2018b). Inversamente, cualquier inconsistencia de estos factores perjudica la fiabilidad y puede darse el caso de que una valoración positiva de la fiabilidad sea el producto de procedimientos o decisiones espurios, que nada tienen que ver con un diseño y aplicación correctos de un sistema de evaluación.

La fiabilidad se refiere, por tanto, a la consistencia de las inferencias extraídas de los resultados de evaluación; a mayor consistencia entre los resultados aportados por cada evaluador más creíble y confiable resulta ser el sistema de evaluación (Moskal y Leydens, 2000; Jonsson y Svingby, 2007). Las divergencias que pueden comprometer esta consistencia pueden darse tanto en las valoraciones generadas por varios evaluadores –lo que en terminología inglesa se conoce como *inter-rater reliability*– o en las que genera un mismo evaluador en su uso del sistema en diferentes ocasiones –lo que en terminología inglesa se conoce como *intra-rater reliability*– (Docktor et al., 2016; Reddy, 2011).

De lo anterior se concluye que el tema de la validez y la fiabilidad en el trabajo con rúbricas es de importancia capital; como debería serlo también en cualquier otro sistema de evaluación alternativo. El asunto es una de las limitaciones que con más frecuencia se asocia al recurso, dado que numerosos estudios demuestran dificultades en conseguir grados aceptables de estas propiedades. La limitación, sin embargo, no es una característica de los sistemas de evaluación basados en rúbricas sino que es más bien una manifestación del reto que supone contemplar los múltiples factores y las múltiples formas de manifestarse que tienen tanto la validez como la fiabilidad, y de las dificultades que entraña la evaluación de constructos complejos. Para Moskal y Leydens (2000), uno de los puntos de partida se encuentra en el propio diseño de la rúbrica, y continúa en sus procesos de aplicación. Prestar atención a las indicaciones que se han ido recopilando a lo largo del artículo puede contribuir notablemente a mejorar los grados de validez y fiabilidad en la utilización de rúbricas.

Conclusión

El diseño de rúbricas no debería considerarse asunto menor en la reflexión en torno al recurso. Una planificación cuidadosa de su diseño –y de su aplicación– es esencial para favorecer su implementación de manera correcta. Este proceso debe afrontarse desde el conocimiento experto del fenómeno que se pretende evaluar y de aquellos aspectos que hemos ido mencionando y que resultan claves para la creación de un diseño efectivo. La adopción de estas recomendaciones favorece la elaboración de rúbricas acordes con un enfoque pedagógico constructivista. La alineación entre los procesos de enseñanza, aprendizaje y evaluación, la negociación colectiva de significados o la orientación de la práctica educativa hacia el estudiante son algunos de los principios esenciales que fundamentan la educación basada en competencias vigente. Su consideración es pues obligada en el diseño de rúbricas; y las rúbricas así diseñadas pueden ser motivos de reflexión excelentes para promover una mejora de los entornos didácticos en coherencia con el modelo educativo. Otro motivo por el que debería prestarse atención al diseño de rúbricas es por el asunto de la validez y la fiabilidad. El diseño de una rúbrica puede ser tanto una amenaza como una oportunidad para estas dos propiedades que definen cualquier sistema de evaluación. Dada la importancia de establecer grados aceptables en ambas, para considerar como útiles y creíbles las inferencias emitidas por un sistema de evaluación, es indispensable prestar atención a cómo se concibe la rúbrica y a cómo se usa.

A la hora de abordar las diferentes fases de la creación de una rúbrica existen dos conceptos fundamentales: equilibrio y consenso. El equilibrio debe regir el establecimiento tanto de los componentes de la rúbrica como del lenguaje a utilizar. La idea es encontrar un balance entre el exceso, que puede convertir la rúbrica en un recurso abrumador y de difícil aplicación, y el defecto que la puede convertir en un instrumento ambiguo que arroje inferencias no válidas o no fiables sobre el fenómeno evaluado. Por otro lado, el consenso es fundamental para poder llegar a un entendimiento compartido de la rúbrica y, por tanto, poder tomar decisiones significativas sobre aquello que se evalúa. Dado que una rúbrica no es auto-explicativa, o mejor dicho, de bien poco sirve si no está perfectamente integrada en el contexto en que se usará y está bien asimilada por quienes la utilizarán, se requieren al menos breves sesiones de formación en que los participantes puedan construir un consenso sobre su significado y alcance, y una práctica compartida sobre su uso.

Una rúbrica no es solo un instrumento de evaluación, sino que es también un recurso que apoya el aprendizaje y la enseñanza, y como todo recurso educativo requiere de tiempo para su creación y aplicación. Esta concepción del recurso es la que debe ayudar a afrontar la inversión de trabajo que comporta su diseño así como otros retos específicos. Mientras perdure la consideración de las rúbricas como meros instrumentos que resuelven procesos de evaluación, es más probable que su diseño no reciba la atención requerida y que, por tanto, el recurso no despliegue todo su potencial.

Agradecimientos

Esta investigación fue financiada por el Ministerio de Economía y Competitividad y los fondos FEDER [HAR2013-46608-R]; el Instituto de Ciencias de la Educación de la Universidad de Barcelona [REDICE16-1420]; el Vicerrectorado de Política Docente y el Programa de Mejora e Innovación Docente de la Universidad de Barcelona [GINDOC-UB/103].

Referencias

- Adedoyin, C.A. (2013). Debate-Proof Grades: Experiences and Challenges of Using a Grading Rubric in a Social Welfare Policy Course. *Journal of Teaching in Social Work*, 33, 196–208.
- Alcón Latorre, M., Menéndez Varela, J. L. y Arbesú García, I. (2017). “Closing the Loop”: rúbricas en la evaluación de programas académicos. *Observar*, 11(2), 115–130.
- Kelly-Riley D. y Allen M. L. (2006). Promoting Undergraduate Critical Thinking in Astro 101 Lab Exercises. *Astronomy Education Review*, 4(2), 10.
- Allen D. y Tanner, K. (2006). Rubrics: Tools for Making Learning Goals and Evaluation Criteria Explicit for Both Teachers and Learners. *CBE - Life Sciences Education*, 5(3), 197–203.
- Anglin, L., Anglin, K., Schumann, P. y Kaliski, J. (2008). Improving the Efficiency and Effectiveness of Grading Through the Use of Computer-Assisted Grading Rubrics. *Decision Sciences Journal of Innovative Education*, 6(1), 51–73.
- Arancibia Aguilera, M. C. (2014). The design of a rubric to evaluate laboratory reports in Astronomy: Academic literacy in the disciplines. *PROFILE Issues in Teachers' Professional Development*, 16(1), 153–165.
- Atkinson, D. y Lim, S. L. (2013). Improving assessment processes in Higher Education: Student and teacher perceptions of the effectiveness of a rubric embedded in a LMS. *Australasian Journal of Educational Technology*, 29(5), 651–666.
- Bekki, J. M., Bernstein, B., Fabert, N., Gildar, N. y Way, A. (2014). Efficacy of an Online Resource for Teaching Interpersonal Problem Solving Skills to Women Graduate Students in Engineering. *Advances in Engineering Education*, 4(2), 1–27.
- Belanger, J., Zou, N., Mills, J. R., Holmes, C. y Oakleaf, M. (2015). Project RAILS: Lessons Learned about Rubric Assessment of Information Literacy Skills. *Portal: Libraries & the Academy*, 15(4), 623–644.
- Bell, A., Mladenovic R. y Price, M. (2013). Students' perceptions of the usefulness of marking guides, grade descriptors and annotated exemplars. *Assessment & Evaluation in Higher Education*, 38(7), 769–788.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347–364.

- Brookhart, S. M. (1999). *The Art and Science of Classroom Assessment: The Missing Pedagogy*. ASHE-ERIC Higher Education Report (Vol. 27, No. 1). Washington DC: The George Washington University, Graduate School of Education and Human Development.
- Chan C. K. Y. (2015). Use of animation in engaging teachers and students in assessment in Hong Kong higher education. *Innovations in Education and Teaching International*, 52(5), 474–484.
- Docktor, J.L., Dornfeld, J., Frodermann, E., Heller, K., Hsu, L., Jackson, K. A., Mason, A., Ryan, Q. X. y Yang, J. (2016). Assessing Student Written Problem Solutions: A Problem-Solving Rubric with Application to Introductory Physics. *Physical Review Physics Education Research*, 12(1), 010130-1–010130-18.
- Eshun, E. F. y Osei-Poku, P. (2013). Design Students Perspectives on Assessment Rubric in Studio-Based Learning. *Journal of University Teaching & Learning Practice*, 10(1), 1–15.
- Etkina, E., Van Heuvelen, A., White-Brahmia, S., Brookes, D. T., Gentile, M., Murthy, S. y Warren, A. (2006). Scientific abilities and their assessment. *Physical Review Special Topics - Physics Education Research*, 2(2), 1–15.
- Franklin, R. y Smith, J. (2015). Practical Assessment on the Run iPads as an Effective Mobile and Paperless Tool in Physical Education and Teaching. *Research in Learning Technology*, 23, 1–19.
- Frazee, J. P. (2001). Charting a Smooth Course for Portal Development. *Educause Quarterly*, 24(3), 42–48.
- Frichlin, L. (2006). *Blueprint for Learning: Constructing College Courses to Facilitate, Assess, and Document Learning*. Herndon (EEUU): Stylus Publishing, LLC.
- Gezie, A., Khaja, K., Chang, V. N., Adamek, M. E. y Johnsen, M. B. (2012). Rubrics as a Tool for Learning and Assessment: What Do Baccalaureate Students Think? *Journal of Teaching in Social Work*, 32(4), 421–437.
- Griffin, M. (2009). What is a rubric? *Assessment Update*, 21(6), 4, 13.
- Hamill, J., Best, G. y Anderson, J. (2015). Developing Student Mentor self-regulation skills through formative feedback: Rubric development phase. *Journal of Peer Learning*, 8, 48–58.
- Hegler, K. L. (enero, 2003). Using General Education Assessment Rubrics to Document Basic Skills and Content Knowledge. En *Annual Meeting of the American Association of Colleges for Teacher Education*. New Orleans, LA. Recuperado de <http://files.eric.ed.gov/fulltext/ED472812.pdf>
- Huba, M. E. y Freed, J. E. (2000). *Learner-centered assessment on college campuses: Shifting the focus from teaching to learning*. Boston: Allyn & Bacon.

- Jonsson, A. y Svingby, G. (2007). The Use of Scoring Rubrics: Reliability, Validity and Educational Consequences. *Educational Research Review*, 2(2), 130–144.
- Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment & Evaluation in Higher Education*, 39(7), 840–852.
- Knight, L. A. (2006). Using rubrics to assess information literacy. *Reference Services Review*, 34(1), 43–55.
- Lewis, R., Berghoff, P. y Pheene, P. (1999). Focusing students: Three approaches for learning through evaluation. *Innovative Higher Education*, 23(3), 181–196.
- Lund, J. (2006). Creating Rubrics for NCATE Reports. *Journal of Physical Education, Recreation, and Dance*, 77(3), 13–17.
- Lundstrom, K., Diekema, A. R., Leary, H., Haderlie, S. y Holliday, W. (2015) Teaching and Learning Information Synthesis: An Intervention and Rubric Based Assessment. *Communications in Information Literacy*, 9(1), 60–82.
- Maxwell, S. (2010). Using Rubrics to Support Graded Assessment in a Competency Based Environment. Occasional Paper. Adelaide: National Centre for Vocational Education Research.
- Mckeown, R. (2011). Using Rubrics to Assess Student Knowledge Related to Sustainability: A Practitioner's View. *Journal of Education for Sustainable Development*, 5(1), 61–74.
- McNamara, J. y Burton, K. (2009). Assessment of Online Discussion Forums for Law Students. *Journal of University Teaching and Learning Practice*, 6(2), 1-14.
- Menéndez Varela, J. L. (2013). Rúbricas para la evaluación de proyectos de Aprendizaje-Servicio en los estudios universitarios de las artes. *Observar*, 7, 5–24.
- Menéndez Varela, J. L. y Gregori Giralt, E. (2016a). Las rúbricas en la valoración de las guías docentes. Reflexiones sobre su impacto y mejora educativa. *Observar*, 10(1), 84–102.
- Menéndez-Varela, J. L. y Gregori-Giralt, E. (2016b). The contribution of rubrics to the validity of performance assessment: a study of the conservation–restoration and design undergraduate degrees. *Assessment & Evaluation in Higher Education*, 41(2), 228-244.
- Menéndez-Varela, J. L. y Gregori-Giralt, E. (2018a). Rubrics for developing students' professional judgement: A study of sustainable assessment in arts education. *Studies in Educational Evaluation*, 58, 70–79.
- Menéndez-Varela, J. L. y Gregori-Giralt, E. (2018b). The reliability and sources of error of using rubrics-based assessment for student projects. *Assessment & Evaluation in Higher Education*, 43(3), 488-499.
- Moni, R. W. y Moni, K. B. (2008). Student Perceptions and Use of an Assessment Rubric for a Group Concept Map in Physiology. *Advances in Physiology Education*, (32)1, 47–54.

- Montgomery, K. (2002). Authentic Tasks and Rubrics: Going beyond Traditional Assessments in College Teaching. *College Teaching*, 50(1), 34–39.
- Moskal, B. M. (2000). Scoring Rubrics: What, When and How? *Practical Assessment, Research & Evaluation*, 7, 22–29.
- Moskal, B. M. y Leydens, J. A. (2000). Scoring Rubrics Development: Validity and Reliability. *Practical Assessment, Research & Evaluation*, 7, 71–81.
- Newell, J. A., Dahm, K. D. y Newell, H. L. (2002). Rubric development and inter-rater reliability issues in assessing learning outcomes. *Chemical Engineering Education*, 36(3), 212–215.
- Oakleaf, M. (2008). Dangers and Opportunities: A Conceptual Map of Information Literacy Assessment Approaches. *Portal: Libraries and the Academy*, 8(3), 233–253.
- Oakleaf, M., Millet, M. S. y Kraus, L. (2011). All Together Now: Getting Faculty, Administrators, and Staff Engaged in Information Literacy Assessment. *Portal: Libraries and the Academy*, 11(3), 831–852.
- Pagano, N., Bernhardt, S. A., Reynolds, D., Williams, M. y McCurrie, M. K. (2008). An inter-institutional model for college writing assessment. *College Composition and Communication*, 60(2), 285–320.
- Petkov, D. y Petkova, O. (2006). Development of Scoring Rubrics for IS Projects as an Assessment Tool. *Issues in Informing Science and Information Technology*, 3, 499–510.
- Raposo Rivas, M., Cebrián de la Serna, M. y Martínez-Figueira, E. (2014). Electronic Rubrics to Assess Competences in ICT Subjects. *European Educational Research Journal*, 13(5), 584–594.
- Reddy, M. Y. y Andrade, H. (2010). A Review of Rubric Use in Higher Education. *Assessment & Evaluation in Higher Education*, 35(4), 435–448.
- Reddy, M. Y. (2011). Design and Development of Rubrics to Improve Assessment Outcomes: A Pilot Study in a Master's Level Business Program in India. *Quality Assurance in Education*, 19(1), 84–104.
- Rezaei, A. R. y Lovorn, M. (2010). Reliability and Validity of Rubrics for Assessment through Writing. *Assessing Writing*, 15(1), 18–39.
- Schlitz, S. A., Connor, M. O., Pang, Y., Stryker, D., Markell, S., Krupp, E. y Redfern, A. K. (2009). Developing a Culture of Assessment through a Faculty Learning Community: A Case Study. *International Journal of Teaching and Learning in Higher Education*, 21(1), 133–147.
- Shaw, G. F. (2014). Introducing Rubrics to Physical Education Teacher Candidates. *Journal of Physical Education, Recreation & Dance*, 85(6), 31–37.

- Simon, M. y Forgette-Giroux, R. (2000). A Rubric for Scoring Postsecondary Academic Skills. *Practical Assessment, Research & Evaluation*, 7, 139–143.
- Slomp, D. (2015). Writing Assessment in Six Lessons--from "American Idol". *Phi Delta Kappan*, 96(5), 62–67.
- Sridharan, B., Leitch, S. y Watty, K. (2015). Evidencing Learning Outcomes: A Multi-Level, Multi-Dimensional Course Alignment Model. *Quality in Higher Education*, 21(2), 171–188.
- Stellmack, M. A., Konheim-Kalkstein, Y.L., Manor, J. E., Massey, A.R. y Schmitz, J.A.P. (2009) An Assessment of Reliability and Validity of a Rubric for Grading APA-Style Introductions, *Teaching of Psychology*, 36(2), 102–107.
- Stevens, D. D. y Levi, A. J. (2005). *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning*. Sterling, VA: Stylus Publishing.
- Suskie, L. (2004). *Assessing student learning: A common sense guide*. Bolton, MA: Anker.
- Thaler, N., Kazemi, E. y Huscher, C. (2009). Developing a Rubric to Assess Student Learning Outcomes Using a Class Assignment. *Teaching of Psychology*, 36(2), 113–116.
- Timmerman, B. E., Strickland, D., Johnson, R. L. y Payne, J. R. (2011). Development of a "Universal" Rubric for Assessing Undergraduates' Scientific Reasoning Skills Using Scientific Writing. *Assessment & Evaluation in Higher Education*, 36(5), 509–547.
- Wiggins, G. (1998). *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco: Jossey-Bass.
- Wilson, V. A. y Onwuegbuzie, A. J. (noviembre, 1999). Improving Achievement and Student Satisfaction through Criteria-Based Evaluation: Checklists and Rubrics in Educational Research Courses. En *Annual Meeting of the Mid-South Educational Research Association*. Point Clear, AL. Recuperado de <http://files.eric.ed.gov/fulltext/ED436569.pdf>
- Wolf, K. y Goodwin, L. (2007). Evaluating and Enhancing Outcomes Assessment Quality in Higher Education Programs. *Metropolitan Universities*, 18(2), 1–13.
- Wolf, K. y Stevens, E. (2007). The Role of Rubrics in Advancing and Assessing Student Learning. *The Journal of Effective Teaching*, 7(1), 3–14.
- Wolf, K., Connelly, M. y Komara, A. (2008). A Tale of Two Rubrics: Improving Teaching and Learning Across the Content Areas through Assessment. *The Journal of Effective Teaching*, 8(1), 21–32.

Para citar este artículo: Alcón Latorre, M. y Menéndez Varela, J. L. (2018). El diseño de rúbricas: algunos aspectos claves. *Observar*, 12, 1–19.