

“Closing the Loop”: rúbricas en la evaluación de programas académicos

Closing the Loop: Rubrics for programme assessment

Mireia Alcón Latorre¹
Universitat de Barcelona
malcon@ub.edu

José Luis Menéndez Varela
Universitat de Barcelona
menendez@ub.edu

Isabel Arbesú García
Universidad Autónoma Metropolitana
isabel.arbesu@gmail.com

Fecha de recepción del artículo: 7 de octubre de 2017

Fecha de aceptación: 31 de octubre de 2017

Resumen

La evaluación de programas con rúbricas es un ámbito de innovación dentro de la investigación educativa. Al enfoque tradicional que prueba la efectividad de los programas a partir del análisis de los resultados de aprendizaje de los estudiantes, se le suma otro que parte de la evaluación directa de los propios programas académicos. La finalidad última, en ambos casos, es la de un cierre del círculo de evaluación (*“closing the loop”*) que lleve a la implementación de mejoras en los mismos. Para ello se requiere la utilización de instrumentos de evaluación válidos y fiables como pueden serlo las rúbricas. El artículo es un breve estado de la cuestión que fundamenta y expone las investigaciones llevadas a cabo en torno al tema.

Palabras clave: programas, rúbricas, evaluación, educación superior.

Abstract

Programme assessment with rubrics is an area of innovation within educational research. Besides the traditional approach that assesses programme effectiveness from the analysis of students' learning outcomes, there is another one based on direct assessment of programmes. The aim, in both cases, is closing the evaluation loop that leads to programme improvement. This requires the use of valid and reliable assessment methods as rubrics can be. The article is a brief state of the art that presents a theoretical framework of the subject.

Keywords: programmes, rubrics, assessment, higher education.

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

Numerosas investigaciones a nivel internacional están generando estudios con el objetivo de probar recursos y metodologías que contribuyan a la actualización de los modelos y prácticas educativas tradicionales (ej: Cooper y Oliver-Hoyo, 2016; Gabaudan, 2016; Gariepy, Stout y Hodge, 2016; Grangier y Weir, 2016; McKeivitt, 2016; Menéndez y Gregori, 2017; Whitlock y Ebrahimi, 2016; Young, James y Noy, 2016). La necesidad de adoptar un nuevo paradigma educativo es ya una opinión respaldada por buena parte de la comunidad científica, un modelo que pasa, entre otros aspectos, por reivindicar un enfoque de la enseñanza orientado hacia el aprendizaje del estudiante y por asegurar la calidad de los sistemas utilizados, una calidad fundamentada en indicadores significativos.

Los resultados generados por estas investigaciones suponen datos relevantes tanto para las autoridades universitarias como para los sistemas de garantía de calidad de las instituciones académicas (Menéndez y Gregori, 2016), ya que aportan información relevante sobre aquello que funciona y aquello que no, a la vez que orientan sobre futuros cambios y mejoras a implementar (lo que se conoce como “*closing the loop*” o cierre del ciclo de evaluación). Los programas académicos deberían ser de los primeros receptores de estas nuevas implementaciones, dado que son la base de las estructuras y sistemas académicos. La evaluación de programas se presenta, por tanto, como un ámbito esencial desde el que abordar la reforma educativa.

La disponibilidad de recursos que permitan llevar a cabo estos procesos de evaluación de una manera válida y fiable supone uno de los factores clave para el correcto desarrollo de los mismos. Las rúbricas, a pesar de ser tradicionalmente utilizadas para la evaluación y desarrollo de competencias, están probando ser también efectivas para la evaluación de los programas. Varios son los estudios que lo demuestran. Algunos parten del análisis de los resultados de aprendizaje de los estudiantes, concibiéndolo como un indicador sobre la efectividad de los programas (ej: Holliday et al., 2015; Ledoux y McHenry, 2006; Schen, 2013; Song, 2006). Otros lo hacen partiendo directamente del análisis del contenido o diseño del propio programa (ej: Cullen y Harris, 2009; Menéndez y Gregori, 2017; Stanny, González y McGowan, 2015), opción más innovadora y que parece ofrecer una visión más clara en cuanto a la valoración de los programas.

El artículo pretende conformar un panorama general en cuanto a la utilización de rúbricas para la evaluación de los programas académicos. Está dividido en dos partes: una primera muy breve que reflexiona sobre el significado y definición de ciertos conceptos clave; otra más extensa que expone el estado de la cuestión y constituye el cuerpo del artículo. Ésta repasa la producción científica existente en torno al tema, analizando tanto sus marcos teóricos como los diferentes enfoques desde los que se están trabajando.

¿Qué es la evaluación de programas?

A la hora de abordar el tema de las rúbricas en la evaluación de programas, parece adecuado detenerse un momento en cuestiones terminológicas, especialmente dado la frecuencia de uso de conceptos como “programa” o “evaluación de programas”. Desde un punto de vista educativo y partiendo de la literatura relacionada encontramos la siguiente definición:

[Un programas es] un conjunto determinado y estructurado de experiencias de aprendizaje que conduce a una calificación. (Venter y Bezuidenhout, 2008, p.1118)

La descripción puede resultar poco precisa pues despliega un abanico que se abre a multitud de sistemas organizados de contenidos, con posibilidad, además, de encontrar contexto en diferentes ámbitos del sistema académico. De esta manera, podríamos hablar de programas tanto para nombrar el plan de estudios de toda una titulación, como la guía docente de una asignatura. Sin embargo, la anterior definición sí puede servir como punto de partida para comprender lo verdaderamente importante de los programas, esto es: su orientación hacia el aprendizaje de los estudiantes. Parece lógico que se definan los programas académicos como compendios de experiencias de aprendizaje, ya que si éstos no contemplaran contenidos, recursos y estructuras para el desarrollo del aprendizaje del estudiante cabría cuestionarse su función. Por tanto, a pesar de que esta orientación hacia el estudiante no es siempre un rasgo evidente en los programas, debería ser un aspecto inherente a ellos. Un vistazo hacia el significado de lo que se conoce como “evaluación de programas”, clarifica todavía más el asunto. De nuevo, la misma referencia aporta una definición que vincula, claramente, los programas con el aprendizaje:

[La evaluación de programas] es la recopilación y análisis sistemático de información para mejorar el aprendizaje del estudiante. (Venter y Bezuidenhout, 2008, p.1116)

Es de suma importancia subrayar la consideración de que los programas tienen sentido en cuanto se conceptualizan y diseñan sin perder de vista ni al estudiante, ni a los objetivos perseguidos en cuanto a su aprendizaje. Llevar a cabo un proceso de evaluación de programas supone, por tanto, valorar la calidad de éstos en relación con la efectividad y/o adecuación a esos propósitos. La finalidad de la evaluación de programas no es otra, entonces, que la mejora del aprendizaje de los estudiantes. Este aspecto es, con frecuencia, olvidado por la propia comunidad educativa, sumida, a menudo, en complejos procesos de evaluación sin objetivo claro o en satisfacer las exigentes demandas de las agencias de acreditación, no siempre orientadas hacia la calidad educativa. Hablar de evaluación de programas es, por tanto, hablar de compromiso con los estudiantes y con el desarrollo de su aprendizaje. Este debería ser un argumento de peso para embarcarse en proyectos o iniciativas que contemplen la evaluación de programas académicos y para justificar su importancia e idoneidad dentro de las medidas para adoptar un nuevo paradigma educativo.

¿Qué investigaciones se están llevando a cabo con las rúbricas?

La evaluación de programas no es un ámbito frecuente en el que se apliquen las rúbricas. A pesar de que estos recursos educativos están haciéndose hueco dentro de los contextos de educación superior, su aplicación habitual es la de la evaluación del desempeño de los estudiantes, concretamente la valoración del mayor o menor desarrollo de determinadas competencias. Este hecho no resulta extraño dado la unánime opinión de la comunidad científica en cuanto a la idoneidad de las rúbricas para este tipo de evaluación. A pesar de ello, sí existe un sector suficientemente representativo de la comunidad científica que está aplicando las rúbricas para la valoración de los programas académicos como para definir una línea de investigación propia. En ella sigue existiendo la tendencia de utilizar las rúbricas bajo el paradigma de la evaluación del desempeño pero se distingue otra que contempla otros enfoques de evaluación.

La primera se destaca por abordar la evaluación de programas desde el análisis de los resultados de aprendizaje de los estudiantes, para, posteriormente, poder emitir valoraciones sobre la efectividad de los programas. La recopilación y análisis de estos resultados se lleva a cabo mediante rúbricas, cuyos datos sirven para respaldar las inferencias emitidas en cuanto a la calidad de los programas. Este tipo de evaluación utiliza un enfoque que puede denominarse como indirecto, ya que no parte del análisis de los propios programas sino de los resultados del desempeño de los estudiantes, entendidos, por tanto, como indicadores de la calidad.

La segunda es más minoritaria en relación con la anterior. Se trata de una línea de investigación que contempla la evaluación de los programas tomando éstos como punto de partida. A diferencia de la otra, ésta utiliza las rúbricas como instrumento de análisis directo de los propios programas, normalmente, de sus diseños y/o contenidos. Mientras que el enfoque indirecto es más fácil de visualizar en el contexto de las aulas, que es donde se produce el contacto con el estudiante y por tanto la recopilación de sus resultados de aprendizaje, el directo se relaciona más con los contextos departamentales o institucionales, que es desde donde deben partir las iniciativas de revisión de los programas, usualmente promovidas por grupos de profesionales expertos comprometidos con la calidad académica.

Evaluación indirecta de los programas

La consideración de una evaluación de los programas partiendo del análisis de los resultados de aprendizaje de los estudiantes, se basa en lo que se conoce como ciclo o circuito de evaluación. La idea que lo fundamenta se asienta en el convencimiento de que la evaluación del desempeño de los estudiantes no debe ser un proceso de evaluación *per se*, esto es, únicamente para valorar o calificar al estudiante. Debe servir, además, para recoger datos que supongan evidencias sobre las que apoyar inferencias relativas a los programas y justificar sus mejoras, en caso de presentarse como necesarias.

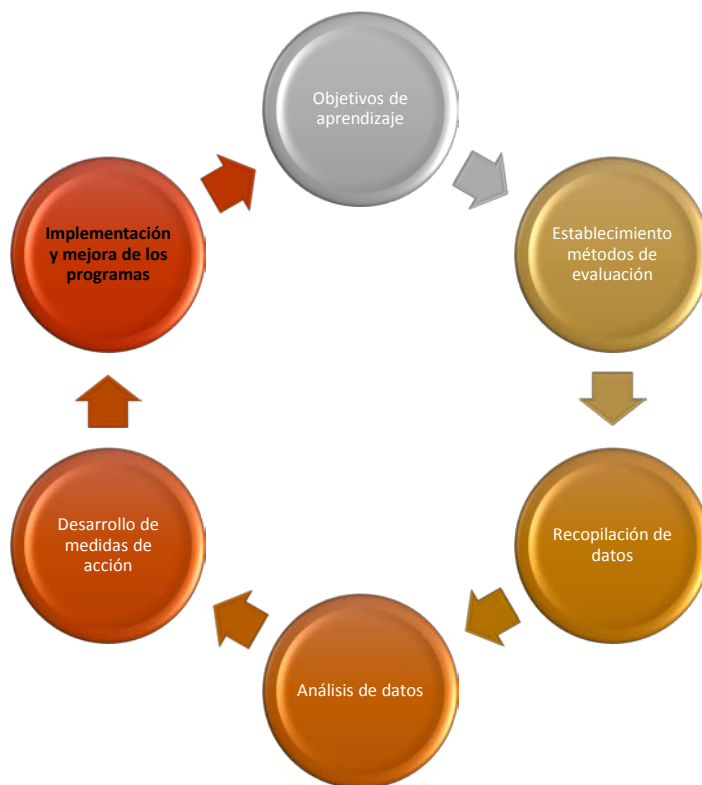


Figura 1. Ciclo de la evaluación. Fuente: original de la autora

Tal y como muestra la figura 1, el ciclo contempla la evaluación como un circuito cerrado que responde o se orienta hacia esa introducción de mejoras en los programas. Estas mejoras deben revertir en el aprendizaje de los estudiantes, de ahí que la trayectoria que marca el circuito continúe en los objetivos de aprendizaje, pues conviene monitorizar las medidas introducidas en relación con los propósitos establecidos. De la claridad en el establecimiento de unos objetivos de aprendizaje depende la solidez del proceso de evaluación, ya que todas las acciones derivadas tienen sentido en cuanto se relacionan con éstos.

Otra de las fases decisivas del ciclo es la del establecimiento de los métodos de evaluación. Seleccionar instrumentos de evaluación válidos y fiables en cuyo diseño participe tanto colectivos de docentes como de estudiantes, es clave para garantizar la efectividad del proceso. Desde las dos últimas décadas, las rúbricas han tomado protagonismo en este sentido, éstas han adquirido valor ya no solo como meros instrumentos medidores, sino como recursos que posibilitan una recopilación sistemática y un análisis de los resultados de aprendizaje del estudiante. Su potencial radica en el establecimiento de unos criterios de evaluación bien definidos y de sus correspondientes descripciones cualitativas, que facilitan el diagnóstico de lo observado y respaldan los juicios de calidad. En este sentido las rúbricas son reconocidas por la literatura como verdaderos “iluminado-

res de la enseñanza y el aprendizaje”² por las oportunidades que brindan para monitorizar y diagnosticar ambos procesos (Baughin, Brod y Page, 2002; Boix, Dawes, Wolfe y Haynes, 2009; Frichlin, 2006; Gezie, Khaja, Chang, Adamek y Johnsen, 2012; Laird, Johnson y Alderman, 2015; Menéndez, 2013; Orzoff, Peinovich y Riedel, 2008; Popham, 1997; Reddy y Andrade, 2010; Reddy, 2011; Rhodes, 2010; Schlitz et al. 2011; Shaw, 2014; Sridharan, Leitch y Watty, 2015; Wolf, Connelly y Komara, 2008; Wolf y Goodwin, 2007). Insertar rúbricas de una manera articulada en el ciclo de evaluación supone un uso modélico de las mismas, pues su integración facilita la alineación entre el desempeño de los estudiantes y los contenidos establecidos en los programas, pudiendo proporcionar una información valiosa para el desarrollo de mejoras pertinentes (Reddy, 2011; Wolf y Goodwin, 2007; Wolf et al., 2008). He aquí la importancia del establecimiento de unos métodos de evaluación válidos que aporten información de calidad y posibiliten diagnósticos precisos.

En efecto, la información resultante de la aplicación de las rúbricas y de su posterior análisis dispone a detectar puntos negros o áreas de mejora y a plantear medidas de acción. Por ello, el ciclo de evaluación establece su continuidad en el punto desde el que iniciamos la explicación: la implementación de mejoras en los programas. Esta fase en la que culminaría el ciclo es la que condensa el sentido del proceso, ya que supone el volcado de la información extraída de los resultados de aprendizaje en los programas académicos. La literatura se refiere a este fenómeno con la expresión anglosajona “*closing the loop*” –que significa, literalmente, “cierre del circuito” o “cierre del bucle”– (Baughin et. al, 2002; Belanger, Zou, Mills, Holmes y Oakleaf, 2015; Diller y Phelps, 2008; Dunbar, Brooks y Kubicka-Miller, 2006; Peach, Mukherjee, Hornyak, 2007; Reddy, 2011; Sridharan et al., 2015; Whitlock y Ebrahimi, 2016; Wolf y Goodwin, 2007). En una concepción holística de la evaluación, este sería el último paso de un proceso que se repetiría de manera cíclica (Wolf y Goodwin, 2007), y que debería tener siempre en el punto de mira el aprendizaje del estudiante.

El término “*feedback*” (retorno, retroalimentación) es a menudo añadido a la expresión “*closing the loop*” (“*closing the feedback loop*”) lo que remarca lo necesario de contar con una información o un “retorno” para poder cerrar de manera efectiva el ciclo. Son muchos los sistemas de evaluación que proporcionan información sobre el desempeño de los estudiantes, pero más importante que la mera existencia o ausencia de un retorno de resultados es la calidad de la información proporcionada (Etkina et al., 2006), ya no solo de cara a empoderar al propio estudiante sino también de cara a validar los procesos de evaluación de programas. La literatura establece una relación directa entre la utilización de rúbricas y la mayor calidad del retorno obtenido sobre el desempeño del estudiante, calidad que se define, especialmente, en términos de utilidad, claridad, capacidad informativa y especificidad (Anglin L., Anglin, Schumann y Kaliski, 2008; Belanger et al., 2015; Bha-

² Esta expresión está adaptada de su versión original inglesa: “Instructional illuminators” (Popham, 1997, 75)

ruthram, 2015; Bird y Yucel, 2013; Cheng y Wang, 2007; Ciorba y Smith, 2009; Etkina et al., 2006; Gabaudan, 2016; Gezie et. al., 2012; Oakleaf, 2008; van Helvoort, 2012). De ahí el valor específico de las rúbricas como recursos idóneos para ser integrados en el proceso de evaluación y facilitar el cierre del ciclo.

Son varios los estudios que utilizan rúbricas para abordar la evaluación de programas desde este enfoque indirecto, esto es, tomando los resultados de aprendizaje de los estudiantes como fuente de información. Su análisis parte del trabajo del aula pero orientan su foco de atención a los objetivos de las titulaciones o áreas de conocimiento bajo influencia de los departamentos y/o facultades, por tanto, su propósito final es la evaluación de las líneas programáticas que éstas establecen. Entre estos estudios, una minoría lleva a cabo el proceso mediante la monitorización del progreso de los estudiantes a lo largo de la titulación (ej: Holliday et al., 2015; Shadle, Brown, Towns y Warner, 2012; Song, 2006), mientras que la mayoría lo hace analizando el grado de cumplimiento de los objetivos establecidos en las titulaciones –a menudo, estándares marcados por las agencias de acreditación– (ej: Cooner, Stevenson y Frederiksen, 2011; Diller y Phelps, 2008; Dunbar et al., 2006; Eppes, Milanovic y Sweitzer, 2012; Good, Osborne y Birchfield, 2012; Jastram, Leebaw y Tompkins, 2014; Ledoux y McHenry, 2006; Lund, 2006; Moseley, Desjean-Perrotta y Utley, 2010; Newel, J.A., Dahm y Newel, 2002; Pagano, Bernhardt, Reynolds, Williams y McCurrie, 2008; Peach et al., 2007; Raudenheimer, 2003; Reed, Levin y Malandra, 2011; Schen, 2013).

De entre los anteriores estudios cabe decir que no son muchos los que informan de un cierre del ciclo de la evaluación mediante la introducción de mejoras (ej: Diller y Phelps, 2008; Holliday et al., 2015; Moseley et al., 2010; Pagano et al., 2008; Peach et al., 2007; Reed et al., 2011). Aunque sí existen otros que, a pesar de no haberse planteado inicialmente la implementación de mejoras en los programas, informan de su puesta en marcha en vista de la necesidad desvelada gracias a la aplicación de las rúbricas (ej: Belanger et al., 2015; Garipey et al., 2016; Kiel, Burclaff y Johnson, 2015; Mills, Bennett, Crawford y Gould, 2009; Whitlock y Ebrahimi, 2016). Estas fuentes dan testimonio de la utilidad de las rúbricas para una recopilación sistemática de información sobre el desempeño de los estudiantes que sirva, a su vez, como evidencia para apoyar los juicios de calidad sobre los programas.

Una vez establecido el panorama general del marco teórico y de la producción científica que conforma esta línea de investigación, es importante añadir una consideración que no debe pasarse por alto. Según lo dicho hasta ahora, podría parecer que existe un imaginario camino recto entre el desempeño de los estudiantes y los contenidos establecidos en los programas. Si bien parece lógico pensar que una baja calidad en los programas tiene –o debería tener– una repercusión negativa en el aprendizaje de los estudiantes, el asunto no parece tan claro si se recorre el camino inverso. Dicho de otra manera: unos resultados de aprendizaje no satisfactorios en el desempeño

del estudiante no tiene porqué implicar una mala calidad de los programas. Y es que no es posible obviar, a este respecto, el contexto educativo en el que se sucede el aprendizaje. Existe una multiplicidad de factores dentro de este contexto que pueden repercutir tanto en el trabajo de los estudiantes cómo en la propia aplicación de los programas, sin que ello comprometa la calidad de éstos. La pertinente consideración de estos factores ampliaría los límites invisibles del ciclo de evaluación, haciéndolo más permeable a otros condicionantes. Es cierto que, los resultados de aprendizaje de los estudiantes pueden servir de indicador de la efectividad de los programas y, por tanto, pueden informar de posibles mejoras a implementar, pero no de ello cabe deducir un sentido unidireccional entre ambos. Otros factores propios del contexto educativo se ponen en marcha, expandiendo a incontables escenarios y circunstancias adyacentes la relación entre resultados de aprendizaje y contenidos de los programas. Tal y como proponen Stanny, González y McGowan (2015) en relación a su estudio, futuras investigaciones podrían ocuparse de casos concretos en los que hayan surgido disparidades entre los programas y su implementación en las aulas, analizando sus posibles causas.

Evaluación directa de los programas

Una alternativa plausible a la evaluación de los programas desde lo que hemos descrito como enfoque indirecto, es el enfoque directo. Tal y como se ha venido comentando hasta ahora, la evaluación directa de los programas no toma como punto de partida los resultados de aprendizaje de los estudiantes, sino los contenidos y/o diseño de los mismos.

A pesar de que el modelo del ciclo de la evaluación está conceptualizado bajo el paradigma de la evaluación del desempeño, el mismo sistema podría adaptarse para la evaluación directa de los programas. El puntal del mismo ya no lo conformarían unos objetivos de aprendizaje sino unos estándares de calidad sobre los que contrastar los resultados de la evaluación de los programas. El resto de fases podrían continuar siendo válidas tal y como quedan establecidas en el ciclo, culminando, de igual manera, en la implementación de mejoras o *closing the loop* según los resultados obtenidos. La unidireccionalidad de este enfoque de evaluación, sin embargo, convierte el cierre del circuito en un desenlace más evidente, pues se entiende que una evaluación directa de los programas tiene como propósito principal la mejora de los mismos. De ahí que no sea habitual tomar este modelo para este otro enfoque de evaluación.

La fase del establecimiento de unos métodos de evaluación supondría, al igual que anteriormente, un paso clave en el proceso. Las rúbricas vuelven a presentarse como recursos aptos para abordar un enfoque directo de evaluación de programas. A pesar de ello, esta línea de trabajo es muy minoritaria, sobrepasada con gran diferencia por la que utiliza las rúbricas bajo el paradigma de la evaluación del desempeño. Las razones por las que el uso de rúbricas no está extendido en este tipo de evaluación no están claramente definidas en la literatura, pero pueden atribuirse a la novedad que supone este tipo de aplicación, que altera el uso habitual que se le atribuye al recurso. En

efecto, la comunidad científica se encuentra sumida en cubrir la creciente necesidad de evaluar por competencias y de testar recursos que prueben ser eficientes para tal efecto. La efectividad de las rúbricas para esta tarea está probando ser satisfactoria, de ahí que se estén generando gran cantidad de estudios probando la aplicación del recurso. Otros ámbitos de uso para las rúbricas son todavía terrenos poco explorados, lo que convierte a la utilización de rúbricas para una evaluación directa de los programas en un método de trabajo innovador.

La utilidad de las rúbricas para este tipo de evaluación viene validada por los numerosos beneficios que se le otorgan al recurso. Los más significativos y que plantean ventajas respecto a los que albergan otros sistemas de evaluación son los siguientes:

- Son sistemas de evaluación en base a criterios o estándares referenciados, que sirven de pautas o referencias a la hora de pensar en conceptos complejos de establecer como el de la calidad educativa.
- Permiten una evaluación tanto cuantitativa como cualitativa, lo que ayuda a emitir diagnósticos más precisos.
- Establecen diversos niveles de consecución de los criterios o estándares establecidos y, por tanto, una imagen más concreta de dónde se está y a dónde se puede llegar.
- Su creación es —o debería ser— fruto del trabajo colaborativo y del consenso de un colectivo profesional comprometido con la calidad educativa.
- Son recursos que se pueden —y deben— adaptar a los contextos educativos concretos en los que van a ser aplicados.

La literatura existente al respecto, a pesar de no ser muy numerosa, supone un incipiente testimonio de la viabilidad del uso de las rúbricas para este tipo de evaluación. Dos grupos de estudios dividen la producción científica: por un lado están aquellos cuyo propósito es la evaluación de las guías docentes, por el otro los que se concentran en la evaluación de las modalidades on-line de los programas, analizando su efectividad y adecuación. La tipología de los estudios de este último grupo no muestra todavía un patrón uniforme, encontrando investigaciones teóricas, experiencias innovadoras y algún estudio cuantitativo puntual. Además, el objeto de análisis se extiende más allá de los propios programas, incluyendo también aspectos prácticos concernientes a las plataformas donde éstos se albergan, por lo que parece una línea de investigación todavía en fase de desarrollo.

Más robustez parece reflejar la línea de investigación que se centra en la evaluación de las guías docentes, que aunque minoritaria, cuenta con estudios sólidos tanto por sus planteamientos como por las metodologías utilizadas. Dos son sus estudios más significativos el de Cullen y Harris (2009) y el más reciente de Stanny et al. (2015).

El estudio de Cullen y Harris (2009) tuvo como propósito evaluar el enfoque educativo de varias guías docentes de diferentes facultades. En concreto analizó el grado en el que las guías se orientaban al estudiante. Para ello, sus promotores crearon una rúbrica cuyos resultados de aplicación se trataron de manera cuantitativa. La rúbrica estaba compuesta por tres categorías: comunidad, poder y control, y evaluación, englobando, cada una de ellas, diferentes criterios o estándares de evaluación. La categoría de comunidad contemplaba aspectos como la accesibilidad del profesor, la alineación del programa con objetivos de aprendizaje y el fomento del trabajo colaborativo. La de poder y control se centraba en aspectos como el rol del profesor y del estudiante, la disponibilidad de recursos externos y la tendencia de las guías docentes. Por último, el de evaluación consideraba aspectos como el de las calificaciones, el retorno proporcionado al estudiante, las pruebas de evaluación, los resultados de aprendizaje y los mecanismos para la revisión de resultados. Según los propios autores y promotores del estudio, la rúbrica probó ser efectiva ya que de su aplicación se identificaron diversos aspectos a trabajar con el personal docente. Dos de los más destacados fueron, por un lado, la necesidad de poner énfasis en la alineación entre los resultados de aprendizaje y las actividades de aprendizaje contempladas en las guías docentes –una alineación a menudo inexistente– y, por el otro, la necesidad de plantear el poder y control del docente de una manera más transversal, esto es, dando lugar a la participación del estudiante. La conclusión general del estudio remarca lo formativo que resultó el proceso de evaluación con rúbricas, ya que el análisis de los resultados generó fructíferas discusiones entre colegas y permitió revisar la propia cultura profesional e idear futuras acciones para el desarrollo de los programas.

El estudio de Stanny et al. (2015) también parte del análisis del contenido de las guías docentes. Su finalidad fue la de analizar cuestiones consideradas importantes respecto a la cultura educativa de la institución, como por ejemplo la estructura empleada en la elaboración de los cursos o el enfoque adoptado en la enseñanza y el aprendizaje. En este caso, partieron de una muestra de guías docentes más extensa que en el anterior (todas las guías docentes de una universidad), lo que comportó un período de evaluación de cinco años de duración dividido en cuatro fases de revisión. Para el proceso de evaluación utilizaron una rúbrica cuyos estándares los conformaron aquellos aspectos entendidos como básicos de una guía docente: los resultados de aprendizaje, la descripción de las actividades de aprendizaje y la modalidad de aprendizaje, entre otros. A diferencia del anterior estudio, éste contemplaba la fiabilidad de la rúbrica, trabajada mediante sesiones formativas con los evaluadores. Según expresaron los propios autores, y a pesar de lo abrumador que manifestaron que fue el proceso, el estudio resultó útil para valorar la calidad educativa de las guías docentes. Los resultados aportados por las rúbricas sirvieron para documentar dos aspectos clave en cuanto al enfoque de enseñanza empleado en la institución académica: por un lado, y al igual que en el estudio de Cullen y Harris (2009), consiguieron determinar el grado de orientación hacia el aprendizaje de las guías docentes; por el otro, obtuvieron valoraciones sobre el uso de estrategias de aprendizaje activo y participativo. Los autores también destacaron la utili-

dad del estudio para guiar futuras acciones, tanto en los propios programas de la institución como en los de otros centros educativos.

Los estudios de Cullen y Harris (2009) y de Stanny et al. (2015) suponen las dos investigaciones más significativas hasta la fecha de entre todo un extenso cuerpo de literatura vinculada con las rúbricas³. Esta línea de investigación es, por tanto, un ámbito poco explorado y conforma un nicho de conocimiento al que parece estar respondiendo la comunidad científica.

Conclusiones

Los programas son la base que asienta y justifica las prácticas que conforman la cultura educativa de las instituciones académicas. Contribuir a la evaluación de programas denota, por tanto, compromiso con la calidad educativa. Los programas deben estar vinculados con los objetivos de aprendizaje establecidos, de manera que éstos sean organizados y razonados de cara a resultar significativos para los estudiantes.

La evaluación de programas, bien por enfoque indirecto o directo, fomenta el desarrollo de la cultura profesional mediante la revisión de cuestiones básicas sobre la enseñanza y el aprendizaje (objetivos, enfoques, metodologías, contenidos, recursos, etc.). Esta información no solo es valiosa para aquellos sistemas encargados de garantizar la calidad de las instituciones, sino también para los propios colectivos académicos, ya que fomenta ejercicios de autoevaluación sobre el propio trabajo. La interiorización de una cultura de la evaluación es un indicativo de excelencia en la práctica educativa, promoviendo la reflexión y actualización de la propia cultura profesional. Sin embargo, la mayoría de los estudios revisados sobre la evaluación de programas con rúbricas no responden tanto a este proceso de interiorización sino más bien a necesidades acreditativas, motivación nada desdeñable siempre y cuando esté orientada hacia la calidad educativa y no responda, únicamente, a meros procedimientos burocráticos.

La producción de investigaciones significativas que prueben sistemas de evaluación eficientes para la evaluación de programas es, por tanto, necesaria para cumplir con estos propósitos. A pesar del desconocimiento en cuanto al potencial educativo de las rúbricas y a pesar de que la comunidad científica está todavía muy centrada en probar su eficacia para la evaluación y desarrollo de competencias, ya están apareciendo estudios que prueban su validez y fiabilidad en cuanto a la evaluación de programas se refiere (ej: Menéndez y Gregori, 2017). Utilizar rúbricas en estos procesos aporta solidez a los procedimientos dada la naturaleza sistemática del recurso, que facilita diagnósticos precisos sobre lo observado. Cabe recordar que las rúbricas deben ser

³ Este artículo es resultado de una investigación más amplia sobre el diseño y aplicación de rúbricas para el que se han revisado más de cuatrocientas fuentes relacionadas.

fruto de procesos rigurosos de diseño, así como del consenso de colectivos especializados y comprometidos con la causa educativa.

Adoptar una práctica de evaluación que tienda a un cierre del ciclo o “*closing the loop*” no es tarea fácil y en muchos casos se convierte en quimera, especialmente debido a las rígidas estructuras institucionales que a menudo dificultan el desarrollo de iniciativas para la mejora. Atendiendo a la literatura, no son muchos los estudios que informan de cambios en los programas después de haber llevado a cabo un proceso de evaluación de los mismos, ya sea a partir de la evaluación de los resultados de aprendizaje de los estudiantes o a través del análisis directo del propio programa. A pesar de lo desalentador del asunto, el compromiso con la propia práctica educativa y la convicción de contribuir a su desarrollo deberían convertirse en la motivación que ayudara a no cejar en el empeño y a seguir generando estudios y experiencias significativas.

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

- 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.
- Baughin, J. A., Brod, E. F. y Page, D. L. (2002). Primary Trait Analysis: A Tool for Classroom-based Assessment. *College Teaching* 50(2), 75–80.
- 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.
- Bird, F. L. y Yucel, R. (2013). Improving Marking Reliability of Scientific Writing with the Developing Understanding of Assessment for Learning Programme. *Assessment & Evaluation in Higher Education* 38(5), 536–553.
- Boix Mansilla, V., Dawes, E., Wolfe, C. y Haynes, C. (2009). Targeted Assessment Rubric: An Empirically Grounded Rubric for Interdisciplinary Writing. *Journal of Higher Education*, 80(3), 334–353.
- Cheng, L. y Wang, X. (2007). Grading, Feedback and Reporting in ESL / EFL Classrooms. *Language Assessment Quarterly*, 4(1), 85–107.

- Ciorba, C. R. y Smith, N. Y. (2009). Measurement of Instrumental and Vocal Undergraduate Performance Juries Using a Multidimensional Assessment Rubric. *Journal of Research in Music Education*, 57(1), 5–15.
- Cooner, D., Stevenson, C. y Frederiksen, H. (2011). Teacher Work Sample Methodology: Displaying Accountability of U.S. Teacher Education Program Effectiveness. *Journal of College Teaching & Learning (TLC)*, 8(10), 17–28.
- Cooper, A. K. y Oliver-Hoyo, M. T. (2016). Argument construction in understanding noncovalent interactions: a comparison of two argumentation frameworks. *Chem. Educ. Res. Pract.*, 17(4), 1006–1018.
- Cullen, R. y Harris, M. (2009). Assessing learner-centredness through course syllabi. *Assessment & Evaluation in Higher Education*, 34(1), 115–125.
- Diller, K. R. y Phelps, S. F. (2008). Learning Outcomes, Portfolios, and Rubrics, Oh My! Authentic Assessment of an Information Literacy Program. *Portal: Libraries and the Academy*, 8(1), 75–89.
- Dunbar, N. E., Brooks, C. F. y Kubicka-Miller, T. (2006). Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. *Innovative Higher Education*, 31(2), 115–128.
- Eppes, T. a., Milanovic, I. y Sweitzer, F. (2012). Towards Liberal Education Assessment In Engineering And Technology Programs. *Journal of College Teaching & Learning (TLC)*, 9(3), 171–178.
- 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.
- Frichlin, L. (2006). *Blueprint for Learning: Constructing College Courses to Facilitate, Assess, and Document Learning*. Herndon (EEUU): Stylus Publishing, LLC.
- Gabaudan, O. (2016). Too soon to fly the coop? Online journaling to support students' learning during their Erasmus study visit. *ReCALL*, 28(2), 123–146.
- Garipey, L. W., Stout, J. A. y Hodge, M. L. (2016). Using Rubrics to Assess Learning in Course-Integrated Library Instruction. *Portal: Libraries and the Academy*, 16(3), 491–509.
- 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.
- Good, J. M., Osborne, K. y Birchfield, K. (2012). Placing data in the hands of discipline-specific decision makers: Campus-wide writing program assessment. *Assessing Writing*, 17(3), 140–149.

- Grainger, P. y Weir, K. (2016). An alternative grading tool for enhancing assessment practice and quality assurance in higher education. *Innovations in Education and Teaching International*, 53(1), 73–83.
- Holliday, W., Dance, B., Davis, E.; Fagerheim, B., Hedrich, A., Lundstrom, K. y Martin, P. (2015). An Information Literacy Snapshot: Authentic Assessment across the Curriculum. *College & Research Libraries*, 76(2), 170–187.
- Jastram, I., Leebaw, D. y Tompkins, H. (2014). Situating Information Literacy Within the Curriculum: Using a Rubric to Shape a Program. *Portal: Libraries and the Academy*, 14(2), 165–186.
- Karen, Y., Kimberley, J. y Sue, N. (2016). Exploration of a reflective practice rubric. *Asia-Pacific Journal of Cooperative Education*, 17(2), 135–147.
- Kiel, S. M., Burclaff, N. y Johnson, C. (2015). Learning by Doing: Developing a Baseline Information Literacy Assessment. *Portal: Libraries and the Academy*, 15(4), 747–766.
- Laird, C., Johnson, D. A. y Alderman, H. (2015). Aligning Assessments for COSMA Accreditation. *Journal of Physical Education, Recreation & Dance*, 86(8), 27–33.
- Ledoux, M. W. y McHenry, N. (2006). Electronic portfolio adoption for teacher education Candidates. *Early Childhood Education Journal*, 34(2), 103–116.
- Lund, J. (2006). Creating Rubrics for NCATE Reports. *Journal of Physical Education, Recreation, and Dance*, 77(3), 13–17.
- McKevitt, C. (2016). Engaging Students with Self-Assessment and Tutor Feedback to Improve Performance and Support Assessment Capacity. *Journal of University Teaching and Learning Practice*, 13(1).
- Menéndez Varela, J.L. y Gregori Giralt, E. (2016). 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. (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., Gregori Giralt, E., Arbesú García, I. (2017). Discusión de una rúbrica para valorar la calidad educativa de las guías docentes en la educación superior. *Observar*, 11(1), 1–24.
- Mills, S., Bennett, B., Crawford, C. B. y Gould, L. (2009). A Model for Integrating Assessment across an Undergraduate Political Science Major. *Assessment Update*, 21(6), 1–2, 14–16.
- Moseley, C., Desjean-Perrotta, B. y Utlely, J. (2010). The draw-an-environment test rubric (DAET-R): Exploring pre-service teachers' mental models of the environment. *Environmental Education Research*, 16(2), 189–208.
- 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.
- Orzoff, J. H., Peinovich, P. E. y Riedel, E. (2008). Graduate Programs: The Wild West of Outcomes Assessment. *Assessment Update*, 20(3), 1–2, 14–16.
- 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.
- Peach, B. E., Mukherjee, A. y Hornyak, M. (2007). Assessing Critical Thinking: A College’s Journey and Lessons Learned. *Journal of Education for Business*, 82(6), 313–320.
- Popham, W. J. (1997). What’s Wrong--and What’s Right--with Rubrics. *Educational Leadership*, 55(2), 72–75.
- Raudenheimer, C. D. (abril, 2003). Faculty Conceptions and Practices of Action Research in the NOVA Program. En Annual Meeting of the American Educational Research Association. American Educational Research Association, Chicago.
- 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.
- Reed, T. E., Levin, J. y Malandra, G. H. (2011). Closing the Assessment Loop by Design. *The Magazine of Higher Learning*, 43(5), 44–52.
- Rhodes, T. (2010). Since We Seem to Agree, Why Are the Outcomes so Difficult to Achieve? *New Directions for Teaching and Learning*, 121, 13–21.
- Schen, M. (2013). A comparison of biology majors written arguments across the curriculum. *Journal of Biological Education*, 47(4), 224–231.
- 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.
- Shadle, S. E., Brown, E. C., Towns, M. H. y Warner, D. L. (2012). A rubric for assessing students’ experimental problem-solving ability. *Journal of Chemical Education*, 89(3).
- Shaw, G. F. (2014). Introducing Rubrics to Physical Education Teacher Candidates. *Journal of Physical Education, Recreation & Dance*, 85(6), 31–37.
- Song, K. H. (2006). A conceptual model of assessing teaching performance and intellectual development of teacher candidates: A pilot study in the US. *Teaching in Higher Education*, 11(2), 175–190.

- 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.
- Stanny, C., Gonzalez, M. y McGowan, B. (2015). Assessing the culture of teaching and learning through a syllabus review. *Assessment and Evaluation in Higher Education*, 40(7), 898–913.
- van Helvoort, A. A. J. (2012). How Adult Students in Information Studies Use a Scoring Rubric for the Development of Their Information Literacy Skills. *Journal of Academic Librarianship*, 38(3), 165–171.
- Venter, P. A. y Bezuidenhout, M. J. (2008). A Mechanism for Programme Evaluation in Higher Education. *South African Journal of Higher Education*, 22(5), 1114–1125.
- Whitlock, B. y Ebrahimi, N. (2016). Beyond the Library: Using Multiple, Mixed Measures Simultaneously in a College-Wide Assessment of Information Literacy. *College & Research Libraries*, 77(2), 236–262.
- 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.
- Wolf, K. y Goodwin, L. (2007). Evaluating and Enhancing Outcomes Assessment Quality in Higher Education Programs. *Metropolitan Universities*, 18(2), 1–13.

Para citar este artículo: 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.