

La participación de los estudiantes como evaluadores

Un estudio en las titulaciones universitarias de las artes

EVA GREGORI GIRALT* | JOSÉ LUIS MENÉNDEZ VARELA**

La implantación de la autoevaluación y la evaluación entre iguales entre estudiantes es todavía limitada en los estudios universitarios, y a ello ha contribuido la existencia de dudas acerca de su validez y fiabilidad. Sistemas de evaluación basados en rúbricas pueden favorecer la eliminación de algunos de estos escollos. En este artículo se analiza la influencia de estos entornos de evaluación con el apoyo de rúbricas en el progreso y en los resultados de aprendizaje obtenidos en la elaboración de proyectos por estudiantes de reciente ingreso en Bellas Artes, Conservación-Restauración e Historia del Arte. Los resultados mostraron diferencias estadísticamente significativas entre los estudiantes que asumieron la responsabilidad de la evaluación y los que no. Por consiguiente, se infiere un impacto positivo de la utilización de rúbricas pero también que ese impacto se incrementa ostensiblemente cuando su utilización se acompaña de entornos de autoevaluación o evaluación entre iguales.

Palabras clave

Autoevaluación
Evaluación por pares
Rúbricas
Educación superior
Arte

The implementation of self-assessment and peer evaluation among students is still limited in university studies, and this has contributed to doubts about their validity and reliability. Evaluation systems based on clear rubrics can favor the elimination of some of these pitfalls. In this paper we analyze the influence of these evaluation environments with the support of rubrics in the progress and learning results obtained in the preparation of projects by students who have recently begun to study Fine Art, Conservation and Restoration, and Art History. The results showed statistically significant differences between the students who took responsibility for the evaluation and those who did not. Consequently, a positive impact of the use of rubrics is inferred, and the fact that this impact is significantly increased when its use is accompanied by self-assessment or peer review.

Keywords

Self assessment
Peer review
Headings
Higher education
Art

Recepción: 25 de enero de 2016 | Aceptación: 20 de abril de 2016

- * Profesora asociada de la Facultad de Bellas Artes de la Universidad de Barcelona (España) y coordinadora junior del Observatorio sobre Didáctica de las Artes (ODAS). Doctora en Historia del Arte y posgraduada en Docencia Universitaria. Líneas de investigación: recursos de aprendizaje y evaluación; uso de las TIC en educación superior y entornos de aprendizaje informal. Publicaciones recientes (2015, en coautoría con J.L. Menéndez), "Validity of the Learning Portfolio: Analysis of portfolio proposal for the university", *Instructional Science*, vol. 43, núm. 1, pp. 1-17; (2015), "La percepción de los estudiantes de Bellas Artes sobre lo aprendido en un entorno de aprendizaje basado en problemas", *Revista Mexicana de Investigación Educativa*, vol. 20, núm. 65, pp. 481-506. CE: gregori@ub.edu
- ** Profesor titular de la Facultad de Bellas Artes de la Universidad de Barcelona (España) y Director del Observatorio sobre Didáctica de las Artes (ODAS). Doctor en Geografía e Historia. Líneas de investigación: evaluación del aprendizaje, evaluación curricular, aprendizaje entre pares y entornos de aprendizaje informal. Publicaciones recientes: (2015, en coautoría con E. Gregori), "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*, vol. 41, núm. 2, pp. 228-244; (2015), "La carpeta de aprendizaje y la reflexión sobre la práctica. Análisis de un modelo para la construcción de profesionales reflexivos en el ámbito de las artes", *Procedia. Social and Behavioral Sciences*, vol. 196, pp. 128-134. CE: menendez@ub.edu

INTRODUCCIÓN¹

La presencia de entornos de autoevaluación o de evaluación entre iguales es un indicador importante del paradigma pedagógico que sitúa al estudiante en el centro del proceso educativo (e.g., Thomas *et al.*, 2011); y también lo es del grado de implantación de entornos didácticos en los que el estudiante adquiere mayor compromiso y responsabilidad en la gestión de su aprendizaje. El asunto del control del sistema de evaluación y la importancia de compartir ese control entre estudiantes y profesores ya fue destacado en estudios que hoy pueden considerarse clásicos (Boud, 2000; Shepherd, 2000); sin embargo, investigaciones recientes como la de Kearney (2013) ejemplifican hasta qué punto su implantación en los estudios universitarios es una realidad todavía incipiente. Esta situación es evidente en el sistema universitario español y en determinadas disciplinas que, como las artísticas, muestran renuencias a la participación de los estudiantes en los procesos de evaluación, y se encuentran en un estadio temprano en la implantación de sistemas de evaluación congruentes con el nuevo paradigma educativo.

El concepto de evaluación auténtica predica una evaluación orientada a la construcción de aprendizajes social y profesionalmente significativos; por ello, un sistema de evaluación que aspire a ser reconocido como “auténtico” debe incorporar actividades de valoración mediante las cuales los estudiantes se familiaricen con los juicios habituales en su entorno social y profesional. La autenticidad exige la adecuación de las actividades de evaluación a los objetivos establecidos, al contexto educativo específico y al perfil del estudiante. Por su parte, la teoría de la evaluación sostenible añade, como requisito, que la evaluación debe promover el aprendizaje a lo largo de la vida

(Boud, 2000; Boud y Falchikov, 2005; Singh y Terry, 2008); en otras palabras, la evaluación debe ponerse al servicio del aprendizaje autónomo, cuyo desarrollo requiere de entornos educativos marcados por una intensa actividad metacognitiva.

Desde esta doble perspectiva, parecería difícil prescindir de actividades de aprendizaje en las que el estudiante se evalúe y evalúe a sus compañeros. Lindblom-Ylanne *et al.* (2006) afirman que la participación del estudiante en los procesos de evaluación es una condición indispensable para impulsar la reflexión sobre los procesos y resultados de aprendizaje y sobre la regulación de las conductas de aprendizaje. Kearney (2013) sostuvo la existencia de una relación directa entre metacognición, prácticas de aprendizaje sostenible y autoevaluación, y evaluación entre iguales. Otros autores defienden los efectos positivos de estas modalidades evaluadoras sobre el aprendizaje autónomo y otras habilidades cognitivas que favorecen el aprendizaje a lo largo de la vida (e.g., Bostock, 2000; Boud, 1990; Falchikov, 2007). Un ejemplo representativo de esta posición es la afirmación de Dochy *et al.* (1999) de la efectividad de la autoevaluación y la evaluación entre iguales y la conveniencia de utilizarlas también en actividades de evaluación con propósitos sumativos. Sin embargo, la implantación de sistemas de evaluación en que los estudiantes participen como evaluadores no está exenta de escollos.

Los estudios sobre la percepción de los estudiantes de estas iniciativas ofrece un primer motivo de reflexión. Naturalmente, hay estudios en los que los estudiantes han ponderado el valor de estas propuestas. Por citar sólo dos ejemplos recientes, en el estudio de Mostert y Snowball (2013), el 58 por ciento de los estudiantes de la muestra valoraron positivamente su utilidad para recibir una realimentación

¹ 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 (REDICE14-1321); el Vicerrectorado de Política Docente y el Programa de Mejora e Innovación Docente de la Universidad de Barcelona (GIDCUB-13/103).

constructiva de los compañeros, identificar errores, impulsar una revisión más profunda del propio trabajo y alcanzar, en consecuencia, mejores calificaciones. También Clarke *et al.* (2013) recogieron opiniones favorables en 61 por ciento de los estudiantes sobre la realimentación obtenida de sus compañeros, la cual fomenta, a juicio de los investigadores, el pensamiento crítico y un enfoque profundo del aprendizaje.

Una parte importante de la experiencia negativa de los estudiantes tras actuar como evaluadores se explica por su falta de familiaridad con una actividad que siempre han conocido como una responsabilidad exclusiva del profesor. A ello se añade el esfuerzo que reclama una actividad compleja que involucra procesos de alto nivel cognitivo. Nordrum *et al.* (2013) detectaron vacilación e incertidumbre en los estudiantes tras participar en un entorno de autoevaluación; estos mismos estudiantes consideraron la experiencia infructuosa, debido a la dificultad de adaptarse a los criterios de evaluación, y frustrante, por la carga de trabajo que exigía. Junto con la misma queja por la carga de trabajo, Hanrahan e Isaacs (2001) y Pope (2001) captaron incomodidad en los estudiantes al recibir las valoraciones de sus compañeros. Esta incomodidad ha sido argumentada en ocasiones como una desconfianza ante juicios provenientes de no expertos (Boud *et al.*, 1999; Nestel *et al.*, 2005) o debida a una descompensación entre el esfuerzo que los estudiantes dedicaron a la evaluación de sus compañeros y la calidad de la realimentación recibida de otros estudiantes (Mostert y Snowball, 2013).

Otro tipo de estudios han destacado problemas específicos de la autoevaluación o la evaluación entre iguales que perjudican la validez y fiabilidad de los sistemas de evaluación en los que se integran. Uno de los temas principales del debate es el del comportamiento de los estudiantes en la práctica evaluadora: si bien algunos análisis han destacado la existencia de correlaciones altas entre

las puntuaciones de los estudiantes y de los profesores (*e.g.*, Falchikov y Goldfinch, 2000; Sluijsmans *et al.*, 1999; Topping, 1998), otros han identificado la tendencia de los estudiantes a inflar las puntuaciones, tanto en la evaluación entre iguales (Bostock, 2000) como en la autoevaluación (Topping, 2003), e incluso han detectado sesgos basados en la amistad, género, raza o idioma (Li y Steckelberg, 2005).

Sin negar el concurso de factores de naturaleza socio-afectiva como los señalados por Li y Steckelberg, no se puede olvidar que el factor principal es que la evaluación comporta procesos de alto nivel cognitivo, máxime cuando se trata de la evaluación de procesos o productos complejos, como es el caso de una educación superior basada en competencias. Que esa complejidad afecta también a los profesores es un asunto ampliamente recogido en una literatura que ha demostrado dificultades en construir visiones consensuadas sobre el objeto, criterios y procedimientos de evaluación, y en conseguir una práctica de la evaluación compartida consistente entre diferentes evaluadores, e incluso entre aplicaciones en diferentes momentos por el mismo evaluador. Si esto ocurre entre expertos que comparten una misma cultura profesional, ¿cómo aspirar a que esto no se produzca entre estudiantes? Por esta razón, se ha defendido la utilidad de sistemas de evaluación basados en rúbricas para reducir los problemas específicos que afectan la autoevaluación y la evaluación entre iguales.

Falchikov (2007) esbozó lo que podrían constituir los componentes clave para una mejora de la capacidad evaluadora de los estudiantes; para la autora, era preciso discutir con los estudiantes evaluaciones que les sirvieran de ejemplo y establecer sistemas de evaluación estructurados —por ejemplo, basados en rúbricas analíticas—, antes de afrontar prácticas evaluadoras en las que los estudiantes negocien las características del sistema de evaluación. Parece lógico suponer que, para promover el desarrollo de la capacidad evaluadora en

los estudiantes, es necesario que el entorno de aprendizaje posibilite la comprensión de los objetivos de aprendizaje, la identificación de criterios de evaluación válidos, y la evaluación consistente de las actividades propuestas. La mejora de los resultados de la autoevaluación, y la evaluación entre iguales, han sido relacionadas con la importancia de que los estudiantes dispongan de criterios específicos e instrucciones útiles (El-Mowafy, 2014; Lindblom-Ylanne *et al.*, 2006), y que se minimicen las diferencias en la interpretación de los criterios entre estudiantes y profesores (Brew, 1999; Orsmond *et al.*, 2000). Para el desarrollo de un juicio profesional es indispensable identificar, comprender y practicar las dimensiones, los criterios y los niveles de consecución; y esto debe hacerse de un modo comprensible para los estudiantes a fin de que puedan orientar y regular su esfuerzo —aprendizaje autónomo—, entender cómo ocurre y discurre su proceso de aprendizaje —metacognición—, y facilitar una reflexión compartida sobre estos asuntos con sus compañeros y con el profesor —realimentación—. Las rúbricas se presentan, entonces, como un instrumento útil para acometer estas cuestiones.

Sin embargo, los beneficios de la utilización de rúbricas en entornos de autoevaluación y evaluación entre iguales no están exentos de dudas. Jonsson y Svingby (2007) concluyeron la escasez de estudios que informaran sobre la autoevaluación o la evaluación entre iguales acompañadas de rúbricas y denunciaron que la poca calidad técnica de los estudios no hacía recomendable la generalización de los resultados. Lindblom-Ylanne *et al.* (2006) y Bouzidi y Jaillet (2009) recogieron, en una revisión de diversos estudios, la existencia de evidencias contradictorias en relación con la validez y la fiabilidad de estos entornos de evaluación, mientras que Cho *et al.* (2006) confirmaron que los problemas de validez y fiabilidad en la evaluación entre iguales aumentan en los estudiantes de reciente ingreso. El-Mowafy (2014) subraya que la utilización de

las rúbricas en su estudio empírico no impidió la existencia de diferencias significativas entre las puntuaciones de los estudiantes, así como entre éstas y las de los profesores, lo que demostraría que estudiantes y profesores todavía interpretan los criterios de evaluación o aplican los procedimientos evaluadores de modo diverso.

En el ámbito específico de los estudios universitarios de las artes, prácticamente no existen estudios sobre la validez de la participación de los estudiantes en sistemas de evaluación basados en rúbricas. Parkes (2010) utilizó una actividad de autoevaluación basada en rúbricas en los estudios de educación musical y concluyó que éstas ayudaron a los estudiantes a mejorar su desempeño. La iniciativa se acompañó de un estudio descriptivo de las percepciones de los estudiantes mediante el análisis de los comentarios que añadían semanalmente en un diario en línea; sin embargo, no incluyó un análisis de la validez de la evaluación de los estudiantes.

Tucker (2013) realizó un estudio sobre la validez y fiabilidad de un sistema de evaluación que incluía una autoevaluación y evaluación entre iguales de estudiantes matriculados en las titulaciones de Arquitectura y Gestión de la Construcción. La investigación consistió en comparar las puntuaciones con las que los estudiantes valoraron las contribuciones de sus compañeros a actividades de diseño realizadas en grupos de trabajo, con las notas medias ponderadas extraídas de las asignaturas cursadas con anterioridad, y con las calificaciones obtenidas en pruebas de evaluación individuales sobre diseño constructivo —estas dos últimas consignadas por los profesores—. Los resultados mostraron que las puntuaciones emitidas por los estudiantes mantenían una débil correlación con las notas medias ponderadas y una ausencia total de correlaciones con las calificaciones obtenidas en diseño. Esto evidenciaría bien un problema de fiabilidad del sistema de evaluación por el desajuste de puntuaciones entre estudiantes

y profesores, bien que los estudiantes estuvieran evaluando otro tipo de habilidades más directamente relacionadas con la gestión de las dinámicas de trabajo en grupo, lo que cuestionaría la validez del sistema de evaluación en forma de una varianza irrelevante de constructo.

Menéndez-Varela y Gregori-Giralt (2015a) examinaron los beneficios de utilizar rúbricas en la evaluación de proyectos de aprendizaje-servicio en las titulaciones de Conservación-Restauración y Diseño. La muestra consistió en 84 estudiantes de reciente ingreso que fueron evaluados por dos profesores y tres estudiantes de posgrado en tres ocasiones. El diseño de la investigación incluyó diversos análisis factoriales, cálculo de la consistencia interna de las puntuaciones y análisis de la varianza. Los resultados evidenciaron un progreso de aprendizaje gracias al cual los estudiantes alcanzaron, al final del semestre, los resultados de aprendizaje establecidos, de lo que se infirió la validez del sistema de evaluación utilizado. Sin embargo, a pesar de que en la evaluación participaron estudiantes, el estudio no se extendió a la práctica de la autoevaluación o a la evaluación entre iguales propiamente dicha.

Estos autores examinaron con posterioridad una actividad de evaluación entre iguales con estudiantes de reciente ingreso en la titulación de Bellas Artes. En este caso los estudiantes utilizaron, junto con los profesores, un mismo sistema de rúbricas para desarrollar sus proyectos y valorar los resultados de los proyectos de sus compañeros (Menéndez-Varela y Gregori-Giralt, 2015b). El estudio mostró que, como ocurrió con los profesores, los estudiantes detectaron un progreso significativo de los proyectos a medida que avanzaba el curso. Además, los análisis de la consistencia interna y los coeficientes de correlación intraclase de las puntuaciones permitieron inferir un uso consistente de las rúbricas y un elevado nivel de concordancia entre los estudiantes.

Por último, en un trabajo reciente, estos mismos autores investigaron la relación existente entre las calificaciones obtenidas en la elaboración de proyectos y en la evaluación entre iguales mediante un sistema de evaluación basado en rúbricas (Menéndez-Varela y Gregori-Giralt, 2015c). La muestra incluyó 195 estudiantes de las titulaciones de Bellas Artes y de Conservación-Restauración. Los resultados mostraron diferencias significativas en las puntuaciones asignadas en los proyectos a lo largo del curso, de manera que se colige un progreso en los resultados de aprendizaje; sin embargo, se hallaron bajas correlaciones entre las puntuaciones de los proyectos y las de la evaluación entre iguales, así como un diferente nivel de compromiso de los estudiantes entre estas dos actividades, a favor de la elaboración de los proyectos. Este hecho fue confirmado al comprobar que aquellos estudiantes que no superaron la actividad de evaluación entre iguales no fue debido a que sus puntuaciones no encajaran en los umbrales de diseño permitidos respecto de la puntuación de los profesores, sino a su escasa participación en la misma. De ahí se infirió la necesidad de fomentar la participación de los estudiantes en la evaluación implicando un mayor número de profesores y asignaturas a fin de normalizar este tipo de actividad y salvar las reticencias comprensibles ante actividades de aprendizaje novedosas.

Acorde con esta línea de investigación, quedaba por estudiar la existencia de evidencias de que las prácticas de autoevaluación y evaluación entre iguales tienen una incidencia positiva en los resultados educativos en un entorno de aprendizaje basado en proyectos. Con esta finalidad, las preguntas de investigación que guiaron el presente estudio fueron las siguientes: 1) ¿cuál fue la contribución de las rúbricas respecto de la elaboración de proyectos?; y 2) ¿qué influencia tuvieron la autoevaluación y la evaluación entre iguales en la construcción de los aprendizajes? En otras palabras, se pretendió estimar hasta qué punto

las rúbricas sirvieron a los estudiantes para comprender los resultados de aprendizaje perseguidos en una actividad nueva para ellos —la elaboración de proyectos de aprendizaje-servicio—, y los criterios que definían la calidad de dichos resultados, así como orientar, en consonancia, su esfuerzo. Asimismo, se pretendió investigar si la participación de los estudiantes en la evaluación incrementaba su toma de conciencia sobre los puntos fuertes y débiles de su trabajo y, por consiguiente, la calidad de su aprendizaje.

CONTEXTO DIDÁCTICO

Este estudio se realizó en asignaturas programadas en los primeros cursos de los grados de Historia del Arte, Bellas Artes y Conservación-Restauración de la Universidad de Barcelona. Las asignaturas se singularizan por el hecho de que su temática —estética y teoría del arte— está poco representada en los diferentes planes de estudio; por esta razón, son calificadas por los estudiantes como “asignaturas teóricas” frente a las asignaturas-taller de procesos artísticos y las asignaturas de Historia del Arte. El reto principal era demostrar la pertinencia y relevancia de las asignaturas en cuestión en el seno de cada una de las especialidades de arte y cuestionar la dicotomía entre asignaturas de carácter teórico y práctico, ya que esto constituye una falacia epistemológica y tiene efectos perniciosos sobre el aprendizaje. Dado que las asignaturas participantes se ubican en los primeros semestres de las titulaciones, se eligió enfocarlas sobre la comprensión crítica de aquello que define a un profesional del arte; esto es, sobre su pericia en la lectura, interpretación y justificación de los fenómenos artísticos con el fin de producirlos, preservarlos, conservarlos y difundirlos. Desde esta perspectiva debía demostrarse la contribución específica de las disciplinas representadas en las asignaturas.

El enfoque de las asignaturas, la organización de las sesiones en el aula y las actividades de aprendizaje se articularon en dos grandes

ejes de trabajo: por un lado, el eje de la alfabetización visual —mejorar la competencia de la lectura e interpretación de un fenómeno artístico en lo que tiene de distintivo— y la familiarización de los estudiantes con los campos de la estética y la teoría del arte. Por el otro, el eje de la construcción de una identidad profesional en cuyo proceso ocupase un lugar destacado la aplicación y evaluación de las narrativas específicas de las asignaturas en la justificación de cualquier manifestación artística. Las actividades de aprendizaje y el trabajo en el aula integraron el trabajo individual y grupal para que los estudiantes tomaran conciencia de la construcción colectiva del conocimiento. Esta finalidad era indispensable en un contexto educativo en donde el aprendizaje es considerado una responsabilidad personal del estudiante.

La primera actividad consistió en una lectura e interpretación de un fenómeno artístico y supuso la aplicación práctica en el aula de las sesiones en que se exponían los contenidos de las asignaturas. Esta actividad se planteó como un estudio de caso en el que el profesor presentaba una imagen y los estudiantes en plenaria la discutían mientras dos estudiantes conducían la sesión de trabajo y resumían las aportaciones de sus compañeros. Al final, el profesor intervenía para subrayar los aspectos más importantes surgidos en la sesión y apuntar las cuestiones que no se hubieran abordado a fin de convertirlas en motivos de reflexión posterior para los estudiantes. Al terminar el semestre, cada estudiante por separado repitió el proceso en un comentario escrito. Esta actividad de lectura pretendía que los estudiantes describieran y organizaran los aspectos formales particulares de la obra; determinaran el tema de la misma e identificaran problemáticas teórico-artísticas o estéticas directamente relacionadas con ésta; y justificaran cómo se concretaban en la imagen comentada.

La segunda actividad planteó a los estudiantes el diseño y planificación de un proyecto de aprendizaje-servicio dirigido a un

colectivo en riesgo de exclusión social en el que se utilizara el arte, en cualquiera de sus manifestaciones, o la actividad plástica. Los estudiantes debían seleccionar el colectivo, documentarse para tener un conocimiento suficiente del mismo y establecer de qué modo se podía incidir en la resolución de alguno de sus problemas mediante el ejercicio profesional propio de la titulación que cursaban. En esta actividad, el reto estaba en el diseño del proyecto y en la reflexión sobre la tarea del profesional: la pregunta “qué es el arte” se transformó en “para qué sirve” y “cómo debe ser utilizado” en el entorno social inmediato.

Se establecieron cuatro requisitos para la elaboración de los proyectos; el primero fue el de la familiaridad: los estudiantes debían tener contacto con el colectivo elegido y la actuación planteada tenía que ser realizable con los recursos generados en el decurso del semestre. Con ello se pretendió reforzar los aprendizajes vinculándolos con su experiencia personal y subrayar la importancia de la resolución de problemas. Si los colectivos seleccionados eran reales, la reflexión sobre la viabilidad de sus proyectos manifestaba el compromiso deontológico de los estudiantes: debían trabajar de un modo realista sobre grupos humanos, contextos y problemas reales. El segundo fue el diseño y la planificación: los proyectos no se ejecutarían en el marco de las asignaturas; únicamente se prepararían para ser puestos en marcha en momentos posteriores. El motivo de este segundo requisito fue doble: por un lado, se quiso eliminar la tensión que comporta realizar una acción de estas características y así centrar la atención de los estudiantes en la reflexión sobre la función social del arte y el papel del profesional; por el otro, se pretendió fomentar una interiorización paulatina de la estructura de un proyecto. Además, y producto de que se trabajaba sobre casos reales, los proyectos sólo se llevarían a cabo una vez que hubieran alcanzado el nivel de desarrollo necesario, lo que coincidía con el final del semestre. El tercer requisito

fue el del grupo: todos los proyectos fueron elaborados en equipos de trabajo de alrededor de cinco miembros con el fin de enriquecer la reflexión y fortalecer el sentido de pertenencia a una profesión. El cuarto requisito fue el de la exposición y defensa pública del proyecto, dada la importancia de las habilidades comunicativas orales en la práctica profesional.

La exposición oral de los proyectos se distribuyó en tres momentos a lo largo de un semestre con un margen de un mes aproximadamente entre cada uno. Las sesiones de exposición se convirtieron en sesiones de intercambio en las que estudiantes ponentes, estudiantes oyentes y profesores discutían sobre los puntos fuertes y débiles de los diferentes proyectos, aportaban nuevas perspectivas y valoraban el propio acto de la exposición. De este modo, cada grupo de trabajo tenía la posibilidad de reflexionar sobre los comentarios recibidos y orientar su trabajo futuro en la mejora de su proyecto. Esta actividad también apuntaba hacia la toma de conciencia de la construcción colectiva del conocimiento y el desarrollo del sentido de pertenencia a una comunidad de expertos.

Tanto la actividad de lectura e interpretación de un fenómeno artístico, como la actividad de proyecto, fueron acompañadas por dos rúbricas que estuvieron a disposición de los estudiantes desde el comienzo de las clases en la plataforma de aprendizaje utilizada en las asignaturas participantes. Con el fin de garantizar la adecuación de las rúbricas a los estudiantes noveles, fueron diseñadas por los profesores en colaboración con estudiantes de posgrado y mantuvieron una estructura idéntica de cuatro niveles de desempeño. Cada nivel incluía y mejoraba las descripciones cualitativas del nivel precedente e incorporaba algún otro aspecto que añadía complejidad al desarrollo del proyecto. Esta progresividad de los criterios según se ascendía en los niveles de consecución se trasladó también a las tres fases en que se distribuyó el desarrollo de los proyectos. A medida que avanzaban

las fases, las rúbricas mantenían los criterios o dimensiones de la fase anterior —y así se hacía posible la solución de las insuficiencias detectadas— y se añadían criterios que recogían otros aspectos fundamentales del desarrollo de los proyectos.

En cada una de las tres fases de exposición oral, los proyectos fueron evaluados por los profesores y también por los estudiantes mediante las rúbricas. Así, en las exposiciones orales cada estudiante por separado tomó nota de lo sucedido y valoró los proyectos en función de las descripciones cualitativas de las rúbricas; estas descripciones ya las habían utilizado previamente como referencia y guía en la elaboración de sus propios proyectos. Desde el punto de vista práctico, esta actividad se llevó a cabo a través de una aplicación en línea en la que cada evaluador seleccionaba el nivel de desempeño alcanzado por cada proyecto en cada criterio de evaluación y le asignaba una puntuación numérica entre un rango de puntuaciones establecido. Es esta participación de los estudiantes en la evaluación de los proyectos lo que constituye el objeto de análisis de este estudio.

De acuerdo con una enseñanza centrada en el estudiante, y con la convicción de que los procesos de aprendizaje y la construcción del conocimiento se incrementan en contextos cooperativos de responsabilidad compartida, se primó la función facilitadora del profesor. El predominio de esta función obligó al profesor a manifestar su condición de experto y como profesional de una disciplina determinada de un modo distinto al tradicional: acentuó su capacidad de generar ocasiones de aprendizaje de interés para los estudiantes, y su capacidad de establecer el andamiaje cognitivo necesario en cada ocasión para promover la resolución de los problemas que los estudiantes iban afrontando. La primera clave de la actividad del profesor se halló en el diseño de un entorno de aprendizaje, que fue el resultado de una reflexión compartida con otros profesores sobre lo que constituye a un

profesional del ámbito de las artes, sobre su contribución a la sociedad y sobre las estrategias didácticas más adecuadas al perfil de sus estudiantes. Fue aquí, y en relación con este último aspecto, que se estimó indispensable la colaboración de estudiantes de posgrado.

La segunda clave fue la realimentación continuada de los estudiantes en las tutorías semanales con cada uno de los grupos de trabajo y en las plenarios de las exposiciones orales. En estas sesiones de trabajo, las rúbricas ayudaron a subrayar la función facilitadora del profesor: el hecho de que aquellas regulaban las agendas de trabajo y los juicios de los estudiantes, pero también los del profesor, concretó en la práctica la idea de un poder académico más distribuido en que se fundaron la discusión sobre la evolución del aprendizaje y las responsabilidades de todos los participantes. Apoyados en las rúbricas, los profesores proyectaron sobre sus estudiantes una identificación de los logros, una interrogación constante sobre los aspectos no consolidados todavía y una crítica constructiva sobre los resultados —nunca sobre los propios estudiantes— que posibilitara el progreso del aprendizaje.

METODOLOGÍA

Muestra

La muestra estuvo compuesta por un total de 358 estudiantes de las titulaciones de grado de Bellas Artes, Conservación-Restauración y de Historia del Arte de la Universidad de Barcelona. Las dos primeras se imparten en la Facultad de Bellas Artes y la tercera en la Facultad de Geografía e Historia. Los estudiantes de la Facultad de Bellas Artes eran de reciente ingreso y habían cursado mayoritariamente la modalidad de Artes en la enseñanza secundaria. Un 79.3 por ciento eran mujeres de 19 años de mediana (21.7 de media) y 60.9 por ciento no tenía ocupaciones laborales. Los estudiantes de Historia del Arte estaban en segundo año y provenían, por lo

general, de la modalidad de Humanidades y Ciencias Sociales. Un 84.1 por ciento eran mujeres de 20 años de mediana (23.4 de media) y 58.5 por ciento no tenía ocupaciones laborales. En todos los casos, era la primera vez que realizaban proyectos de aprendizaje-servicio.

Diseño metodológico

De acuerdo con la clasificación de Montero y León (2007), se realizó un cuasi experimento, sólo post, con grupos duplicados y cuasi control simultáneo. Los grupos experimentales estuvieron compuestos por 182 estudiantes noveles matriculados en cuatro grupos-aula de la Facultad de Bellas Artes que realizaron la exposición y defensa pública de proyectos de aprendizaje-servicio y participaron en su evaluación mediante rúbricas. La semejanza de estos grupos experimentales radica en que compartieron un perfil de estudiante similar, idéntico entorno didáctico y los mismos profesores. Los grupos de control reunieron 176 estudiantes de la titulación de Historia del Arte inscritos también en cuatro grupos-aula que no participaron en el proceso de evaluación.

Se trabajó con las calificaciones que estos estudiantes habían obtenido en las tres fases de evaluación de los proyectos. La preparación de los datos implicó tres matrices diferentes: la relativa a las calificaciones finales de la asignatura de cada grupo-aula se utilizó para examinar el rendimiento académico de

cada uno de los grupos-aula por separado y en conjunto, y comprobar si la distribución de sus calificaciones se acercaba a una distribución normal. Sobre esta matriz se calcularon los porcentajes de suspensos, aprobados, notables y sobresalientes por grupo-aula y por grupo experimental y de control. En una segunda matriz se reunieron las calificaciones finales de los proyectos ordenadas por grupo experimental y grupo de control. Dicha matriz sirvió para estimar la varianza (ANOVA) de las calificaciones finales de los proyectos de cada uno de estos dos grupos. Por último, la tercera matriz recogió las puntuaciones que obtuvieron los proyectos del grupo experimental y del grupo de control por separado en las tres fases de evaluación. Con estos datos se examinó, mediante ANOVAS, si había diferencias estadísticamente significativas entre las puntuaciones. Con el fin de comprobar los resultados, todos los análisis se realizaron dos veces con dos paquetes estadísticos diferentes.

RESULTADOS

En la Tabla 1 se presentan los resultados académicos (en porcentaje) de los ocho grupos-aula que participaron en el estudio. La Tabla 2 resume estos resultados tras agrupar los datos en un único grupo experimental y un único grupo de control.

Tabla 1. Distribución porcentual de las calificaciones finales por grupo-aula

Grupos	Suspense (%)	Aprobado (%)	Notable (%)	Sobresaliente (%)
Grupo experimental 1	2.2	42.2	48.9	6.7
Grupo experimental 2	2.3	56.8	40.9	0
Grupo experimental 3	8.7	63	21.7	6.5
Grupo experimental 4	0	62.2	35.6	2.2
Grupo de control 1	5.6	35.2	48.1	11.1
Grupo de control 2	34.3	51.4	14.3	0
Grupo de control 3	15.6	42.2	42.2	0
Grupo de control 4	32.4	43.2	21.6	2.7

Fuente: elaboración propia.

Tabla 2. Distribución porcentual de las calificaciones finales por grupo de control y experimental

Grupos	Suspenso (%)	Aprobado (%)	Notable (%)	Sobresaliente (%)
Grupo experimental	3.3	56.1	36.7	3.9
Grupo de control	19.9	42.1	33.9	4.1

Fuente: elaboración propia.

Las Tablas 1 y 2 reflejan una tendencia similar en el rendimiento académico de todos los participantes. La distribución se aproxima a la normal, con un ligero sesgo hacia la derecha. De ello se puede inferir que la propuesta didáctica se ajustó al perfil del estudiante al que iba dirigido. Las tasas de superación de las asignaturas fueron elevadas: 96.7 por ciento en el caso del grupo experimental, y 80.1 por ciento en el caso del grupo de control. No obstante, la Tabla 2 muestra que fue en esta tasa en donde se halla la diferencia entre el grupo experimental y el de control: mientras el grupo experimental presenta muy pocos suspensos (un 3.3 por ciento), el de control arroja casi

un 20 por ciento. Todos los porcentajes de aprobados y notables del grupo experimental superan los del grupo de control y la de sobresalientes apenas difiere en un 0.2 por ciento. También se observa que cuanto mayor es la nota final (notable o sobresaliente) menor es la diferencia entre los dos grupos. Si se analizan estas calificaciones a partir de los descriptores recogidos en las rúbricas, se concluye que más de la mitad de los estudiantes se situó en el tercero de los cuatro niveles de desempeño.

La Tabla 3 resume la estimación de la varianza de las calificaciones obtenidas en la actividad de proyecto por el grupo experimental y el grupo de control.

Tabla 3. ANOVA del proyecto del grupo experimental y del grupo de control

	Suma de cuadrados	gl	Media cuadrática	F	Sig. (p<0.05)
Inter-grupos	24.387	1	24.387	13.135	0
Intra-grupos	660.992	356	1.857		
Total	685.379	357			

Fuente: elaboración propia.

Con un valor máximo explicado por el azar de 2.6, y una fiabilidad del 95 por ciento, se pudo rechazar la hipótesis nula y aceptar que existieron diferencias significativas entre los dos grupos. La media del grupo experimental fue de 7.25, mientras que la del grupo de control fue de 6.73. Se comprueba así que aquellos estudiantes que evaluaron los proyectos de los compañeros obtuvieron

mejores calificaciones en esta actividad. Quedaba por investigar si esta diferencia de promedios a favor del grupo experimental reflejaba también un progreso en las calificaciones a lo largo del semestre. Para ello, se llevó a cabo un nuevo análisis de la varianza de las puntuaciones obtenidas en cada fase por dicho grupo. El resultado se muestra en la Tabla 4.

Tabla 4. ANOVA del proyecto por fases del grupo experimental

	Suma de cuadrados	gl	Media cuadrática	F	Sig. (p<0.05)
Inter-grupos	614.176	2	307.088	192.3	0
Intra-grupos	867.130	543	1.597		
Total	1481.306	545			

Fuente: elaboración propia.

También en esta ocasión, el valor de la F superó al valor máximo explicado por el azar y permitió asegurar que, con un 95 por ciento de fiabilidad, había diferencias estadísticamente significativas entre las tres fases del grupo experimental; en este orden la mejor fue la tercera, con un promedio de 7.25, seguida de la segunda, con una media de 5.60, y de la primera, con un promedio de 4.68. Los estudiantes pasaron de una calificación media

de suspenso al empezar el curso, a una calificación media de notable al terminarlo. La evolución de dichas calificaciones fue positiva durante el semestre, lo que puede interpretarse como el reflejo de un progreso en el aprendizaje a partir de una actividad —el proyecto de aprendizaje-servicio— que les supuso un auténtico desafío cognitivo al principio del curso. La Tabla 5 recoge este mismo análisis en el caso del grupo de control.

Tabla 5. ANOVA del proyecto por fases del grupo de control

	Suma de cuadrados	gl	Media cuadrática	F	Sig. (p<0.05)
Inter-grupos	7.193	2	3.596	2.41	0.09
Intra-grupos	782.272	525	1.490		
Total	789.465	527			

Fuente: elaboración propia.

Con un 95 por ciento de fiabilidad y 2.6 como valor máximo permitido, en el grupo de control hubo que aceptar la hipótesis nula y entender que las diferencias entre fases fueron debidas al azar. La media de la fase 3 (6.73) fue superior a la de la fase 2 (6.60) y a la de la fase 1 (6.44), pero esta ordenación no pudo ser considerada significativa. Estos datos permiten extraer dos conclusiones: primero, que el proyecto de aprendizaje-servicio no les supuso un reto cognitivo tan importante como en el caso de sus compañeros de la Facultad de Bellas Artes, que conformaron el grupo experimental. La calificación media de 6.44 de la primera fase estuvo sensiblemente por encima del 4.68 de estos últimos. Dado que los estudiantes de Historia del Arte se hallaban

sólo un curso por encima, que el tipo de asignaturas estaba infrarrepresentado en las dos facultades, y que tampoco ellos habían realizado antes proyectos de aprendizaje-servicio, la razón de esta diferente puntuación en la primera fase hay que buscarla en la especialidad cursada en la enseñanza secundaria.

La segunda conclusión es que los estudiantes del grupo de control se movieron a lo largo del curso en un umbral muy parecido de puntuaciones, de modo que no puede sostenerse una mejora neta en los resultados de aprendizaje en la tercera fase. A pesar de que debieron de haber desarrollado habilidades cognitivas mejor adaptadas a un aprendizaje basado en proyectos en su paso por la enseñanza secundaria, esta ventaja no se concretó

en un aprendizaje de más calidad al final del curso; incluso fue al contrario, ya que su calificación media en la tercera fase (6.73) se situó por debajo de los estudiantes de Bellas Artes (7.25), quienes revirtieron así sus peores resultados de la primera fase. Llegados a este punto, el factor determinante debió ser la participación de los estudiantes del grupo experimental en la evaluación de los proyectos.

DISCUSIÓN

Resultados académicos en la elaboración del proyecto

Se estimó indispensable que los estudiantes tuvieran, desde el inicio del curso, un sistema de rúbricas como referencia para entender los objetivos de aprendizaje y las expectativas de los profesores al respecto (Hattie y Timperley, 2007), para conseguir una *ratio* razonable entre esfuerzo y resultados (Dunbar *et al.*, 2006), y para mejorar la comunicación con sus compañeros y profesores (Baughin *et al.*, 2002) y establecer las bases de una reflexión compartida (Maxwell, 2010). Esto último debía concretarse en las agendas de trabajo de las reuniones semanales con los profesores y en la comprensión de la calidad de los proyectos de los otros grupos de trabajo.

Las rúbricas no se diseñaron como un conjunto de directrices de aplicación mecánica, sino como documentos que recogían algunos aspectos centrales de la estructura de un proyecto y de la exposición oral. De acuerdo con otros autores (Andrade, 2005; Reynolds *et al.*, 2009; Stupans *et al.*, 2013), las rúbricas fueron un recurso para centrar el diálogo de los participantes y ayudar a clarificar los argumentos con el fin de lograr una realimentación más efectiva. Vinculadas a las rúbricas, las puntuaciones de cada proyecto fueron una representación del grado de consolidación de los aprendizajes dentro de cada nivel de desempeño.

Los porcentajes de las tasas de superación de las asignaturas son un primer indicador de los resultados de aprendizaje alcanzados por

los estudiantes al final del semestre. La distribución normal de las puntuaciones —con un ligero sesgo hacia la derecha— significa que la propuesta didáctica encajó con el contexto y alcanzó un equilibrio razonable entre esfuerzo y logro educativo. Las puntuaciones medias, sobre una escala del 0 al 10, del grupo experimental (7.25) y del grupo de control (6.73), concretan más la calidad de estos aprendizajes; sin embargo, cuando se relacionan con las descripciones cualitativas de las rúbricas, el volumen y calidad de la información se incrementa notablemente; y esto tanto para estudiantes y profesores, como para otros agentes externos interesados.

En cuanto a la definición del proyecto, el nivel de consecución inferior recogía las razones que explicaban la inadecuación del trabajo de los estudiantes a la estructura fundamental de un proyecto de aprendizaje-servicio y a su exposición pública según los criterios de adecuación al acto público, elocución, gramática, contacto visual y adecuación al tiempo disponible. El segundo nivel reflejaba una adecuación del proyecto al colectivo en riesgo de exclusión social y al contexto en que debía llevarse a cabo. El tercer nivel añadía la necesidad de que el proyecto fuera viable, lo que suponía que el grupo de estudiantes debía analizar el proyecto también en relación con las propias capacidades del grupo. Por último, el nivel superior requería que la labor del grupo fuera indispensable para la realización del proyecto, lo que manifestaba una reflexión del estudiante sobre cómo encarnar el profesional de su disciplina. Los dos niveles inferiores implicaban calificaciones inferiores al aprobado, de manera que para superar la asignatura era necesario elaborar un proyecto orientado a la resolución de algún problema bien identificado de un colectivo en riesgo de exclusión social, y que dicho proyecto fuera específico de un equipo de profesionales del ámbito de las artes.

Con idéntico nivel de exigencia, era necesario fundamentar el proyecto en información sobre el colectivo y el contexto de ejecución

extraída de fuentes acreditadas. Además, se debía justificar la pertinencia y la contribución particular del proyecto mediante un estudio comparativo con otros proyectos similares ya concluidos y procedentes de diversos orígenes. Finalmente, la descripción y discusión de las actividades que articulaban el proyecto debían hacerse con detalle en lo referente al tipo de materiales, equipamientos y actuaciones necesarias, y debían incluir una previsión de los posibles problemas que surgirían en su aplicación y las soluciones más razonables. En este caso, la máxima puntuación se alcanzaba con la explicación de las estrategias para lograr la intervención de todos los participantes en el seguimiento del proyecto, su valoración final y las propuestas de revisión.

Las puntuaciones medias antedichas significan que los estudiantes se situaron mayoritariamente en la parte superior del tercer nivel, lo que ilustra la notable calidad de los aprendizajes obtenidos. Estas mismas puntuaciones evidencian que las rúbricas ayudaron a los estudiantes en la construcción de los aprendizajes de las asignaturas, máxime ante el hecho de que la elaboración de proyectos de aprendizaje-servicio y su exposición y defensa pública movilizan procesos cognitivos de alto nivel. Que el grupo de control obtuviera en la primera fase un 6.44 de media, cuando nunca antes habían afrontado el desafío de este tipo de proyectos, podría indicar la utilidad de las rúbricas para mejorar la calidad de la realimentación de los profesores. Posiblemente esto fue así porque focalizaron los comentarios en las actividades que ocuparon a los estudiantes en cada momento del curso (Venning y Buisman-Pijlman, 2013) y les sirvieron para planificar el trabajo de las fases sucesivas (Beaumont *et al.*, 2011). El grupo de control demuestra que, con independencia de su participación en la autoevaluación y en la evaluación entre iguales, todos los estudiantes comprendieron los objetivos de aprendizaje y supieron aplicar los criterios profesionales sobre los que se juzgó su trabajo.

La participación de los estudiantes en la evaluación

Sin embargo, lo anterior no fue garantía de un progreso sostenido del aprendizaje a lo largo del curso, de lo que se desprende que la utilización de rúbricas es una condición necesaria, pero no suficiente, para afirmar la validez de un sistema de evaluación. Aunque las medias del grupo de control se incrementaron a lo largo del semestre, el análisis de la varianza mostró que no existieron diferencias estadísticamente significativas entre las tres fases de evaluación. Todavía más rotundo fue el hecho de que la puntuación media de la primera fase de evaluación del grupo experimental se quedó en un 4.68: en su caso, no bastaron las rúbricas ni la realimentación de los profesores, al menos en la primera fase. De todo lo anterior se concluye que la validez de un sistema de evaluación es el resultado de una confluencia positiva de factores en un contexto determinado; confluencia en la que muy difícilmente puede determinarse el componente con mayor incidencia.

Los análisis de la varianza mostraron diferencias significativas en las puntuaciones entre el grupo experimental y el grupo de control; también revelaron que, a diferencia del grupo de control, las puntuaciones del grupo experimental evolucionaron positivamente a medida que se sucedían las fases de evaluación, y lo hicieron de modo ostensible. Si los estudiantes que compusieron el grupo experimental provenían de una educación secundaria menos adaptada a las actividades de aprendizaje propuestas, el entorno de autoevaluación y evaluación entre iguales fue el elemento diferencial para superar ese déficit inicial y, además, imprimir un mayor progreso y mejores resultados de aprendizaje.

En el estudio se cumplió la advertencia de Sadler (2009) de que el simple acceso a los criterios de evaluación no acaba de funcionar, ni aun reconociendo que las rúbricas mejoraron los procesos de realimentación. La propuesta de entornos didácticos que incluyan la

participación de los estudiantes en la evaluación es indispensable desde la óptica de una educación centrada en el estudiante con el objetivo de impulsar el aprendizaje autónomo como fundamento de un aprendizaje continuado a lo largo de la vida. De acuerdo con Thomas *et al.* (2011), esta experiencia evaluadora se proyecta sobre el aprendizaje futuro en la medida en que impulsa la reflexión del estudiante sobre la calidad de su trabajo. Por consiguiente, este enfoque implica entender la participación de los estudiantes en la evaluación prioritariamente como una actividad de aprendizaje (Clarke *et al.*, 2013; Lindblom-Ylance *et al.*, 2006) encaminada a la construcción del juicio profesional. Sin embargo, aunque los introduzca en los juicios profesionales, no debe olvidarse que esta responsabilidad evaluadora de los estudiantes genera un aprendizaje que, en primera instancia, se refiere al contexto educativo; esto es, aclara qué es lo prioritario y cómo abordarlo con garantías de éxito. Esto refuerza la afirmación de Young (2013) de que la contextualización académica de los criterios utilizados es esencial para alcanzar correlaciones más fuertes entre las puntuaciones de estudiantes y profesores.

La participación de los estudiantes en la evaluación no sólo supuso otra modalidad de realimentación para los grupos evaluados diferente a la recibida de los profesores; también implicó, para los estudiantes evaluadores, otro enfoque al principal problema en cuestión: la elaboración de sus propios proyectos. En efecto, la evaluación entre iguales supuso un espacio de reflexión sobre asuntos centrales que incumbían a los proyectos, pero este espacio tenía la ventaja de interponer una distancia respecto del trabajo del estudiante evaluador en su proyecto, ya que el objeto de análisis era el de sus compañeros. No está suficientemente estudiado el valor de la serendipia en el contexto educativo y sus posibles conexiones con procesos cognitivos complejos como la creatividad, que tanta importancia tiene en el ámbito de las artes; tampoco es un

tema directamente relacionado con la investigación expuesta en este artículo. No obstante, es oportuno insistir en la conceptualización del error como una oportunidad de aprender, y en la conveniencia de que el estudiante asimile esta circunstancia, dada la tendencia a interpretar el error como un menoscabo también personal que perjudica gravemente su autoestima y la confianza entre profesores y estudiantes. Sin fomentar una tolerancia al error en los estudiantes es imposible impulsar habilidades emprendedoras; y sin esa misma tolerancia, no acontece una reflexión serena sobre el propio trabajo que se proyecte en una mejora significativa del aprendizaje. Aquí el análisis y la toma de conciencia del error en los compañeros alcanza sus mejores resultados: identifica fuentes de error comunes y, en consecuencia, despersonaliza el error y lo hace susceptible de un debate compartido y constructivo. Por ello, junto con otras publicaciones (El-Mowafy, 2014; Mostert y Snowball, 2013), esta investigación refuerza la incidencia que tiene el aprender de los errores y aciertos ajenos sobre la capacidad de resolver problemas. De ello se desprende la importancia de que la autoevaluación sea precedida, o como mínimo acompañada, de un entorno de evaluación entre iguales (véase Boud *et al.*, 2001).

La participación en una evaluación basada en rúbricas les sirvió también para promover un aprendizaje autónomo; esto es, para pertrecharlos con estrategias que pueden utilizar independientemente de los profesores y de sus compañeros, así como ser transferidas a otros contextos, dentro y fuera de la institución educativa. Evaluar permite a los estudiantes construir un aprendizaje más amplio y profundo sobre los criterios de calidad, los procesos de revisión y enjuiciamiento de los resultados, los diferentes estilos de evaluación y los problemas más habituales en la aplicación de los procedimientos; en suma, aspectos esenciales de lo que Smith *et al.* (2013) denominaron “*assessment literacy*”. O’Donovan *et*

al. (2004) afirman que los estudiantes necesitan la práctica del juicio para adquirir paulatinamente el conocimiento tácito que le resulta consustancial y así estar en mejor disposición para participar en una reflexión colectiva sobre lo que significa un juicio profesional, sobre sus procesos y sus efectos. Esta actividad de enjuiciamiento debe ser introducida por actividades bien estructuradas en las que las rúbricas pueden ocupar un lugar relevante porque mejoran la información sobre los criterios de evaluación establecidos.

En el caso de nuestra propuesta, esta práctica de la evaluación entendida como aprendizaje se concretó en un juicio de los estudiantes que no computó en la calificación de la nota de los compañeros. No obstante, las puntuaciones de los estudiantes sí se compararon con las de los profesores, se convirtieron en objeto de análisis y reflexión, e incidieron en la calificación de los estudiantes evaluadores. De este modo, se minimizaron las renuencias de los estudiantes a recibir comentarios de compañeros por considerarlos poco capacitados o poco dispuestos a asumir adecuadamente ese compromiso (Boud *et al.*, 1999; Horstmannshof y Brownie, 2013; Mostert y Snowball, 2013; Nestel *et al.*, 2005; Parkes, 2010), sin que ello fuera en menoscabo del reconocimiento de la responsabilidad y del valor de la práctica evaluadora de todos los participantes. Al mismo tiempo, la decisión de integrar la autoevaluación y la evaluación entre iguales en las sesiones de exposición oral de los proyectos —por lo tanto, como actividad desarrollada casi totalmente en el aula— evitó incrementar en exceso la carga de trabajo que los estudiantes atribuyen a esta actividad; se neutralizó así otro escollo detectado en investigaciones anteriores (Hanrahan e Isaacs, 2001; Nordrum *et al.*, 2013; Pope, 2001).

En este estudio se ha intentado evitar la poca calidad técnica que Jonsson y Svingby (2007) advirtieron en las investigaciones sobre las prácticas de autoevaluación y evaluación

entre iguales basadas en rúbricas. La muestra de 358 estudiantes y su distribución equilibrada entre el grupo experimental y el de control (182 y 176 respectivamente) es digna de consideración, sobre todo si se tiene en cuenta que el grupo experimental estuvo constituido por estudiantes de reciente ingreso y que el volumen total de la matrícula de primer curso en la Facultad de Bellas Artes no supera los 350 estudiantes en el conjunto de las dos titulaciones analizadas de esta facultad. Se cuidó también la robustez del diseño metodológico mediante un cuasi experimento con grupos duplicados y cuasi control simultáneo. Por las condiciones naturales en las que se trabajó no fue posible un mayor control de los grupos. Se desestimó un cuasi experimento pre-post porque estaba acreditado que ninguno de los estudiantes había elaborado nunca un tipo de proyecto como el propuesto, no había sido evaluador en ningún tipo de pruebas de evaluación, ni había tenido conocimiento de lo que era una rúbrica.

Este trabajo pretende sumarse a aquellos otros que han defendido una influencia positiva sobre el aprendizaje de compartir con los estudiantes la responsabilidad de la evaluación y la conveniencia de que esta intervención se lleve a cabo con el apoyo de rúbricas (e.g. El-Mowafy, 2014; Falchikov, 2007; Lindblom-Ylance *et al.*, 2006). Sin embargo, hay asuntos importantes que no se han resuelto y determinan líneas de investigación que deberían llevarse a cabo en el futuro: por un lado, está el examen de la corresponsabilidad entre profesores y estudiantes en la evaluación a partir del cálculo de algún coeficiente de consistencia interna y del grado de acuerdo entre ambas tipologías de evaluadores. Por otro, sería útil despejar las fuentes de error que afectan a estos entornos educativos y determinar su impacto relativo a fin de establecer las estrategias oportunas. En este caso, se impone un estudio de fiabilidad desde la teoría de la generalizabilidad.

CONCLUSIONES

Smith *et al.* (2013) destacan la importancia de promover en los estudiantes de reciente ingreso su integración en la cultura académica de las titulaciones que cursan y una mayor responsabilidad sobre su propio aprendizaje. La consecución de ambos objetivos es más fácil si estos estudiantes están en disposición de juzgar su propio trabajo, y para ello es imprescindible su compromiso con la práctica de la autoevaluación y evaluación entre iguales, a pesar de la existencia de dificultades específicas en los estudiantes de primer año (Cho *et al.*, 2006). Nuestro estudio corroboró los beneficios educativos —en estudiantes de reciente ingreso— de participar como evaluadores. Estos estudiantes noveles obtuvieron mejores resultados académicos y un mayor progreso que estudiantes más avanzados, y con un bagaje cognitivo más adecuado a las actividades de aprendizaje propuestas que sólo utilizaron las rúbricas como referencia y guía de su trabajo. Los dos grupos mostraron

una distribución normal de las calificaciones y se puede argumentar, entonces, la adecuación del entorno didáctico a los objetivos de aprendizaje y al perfil de los estudiantes. La única diferencia entre ellos fue que la práctica de la autoevaluación y de la evaluación entre iguales permitió a los estudiantes noveles introducirse mejor en la cultura profesional de su especialidad. Para que esto sea posible es necesario cumplir con la premisa de un enfoque holístico sobre la evaluación en el que se consideren conjuntamente las aportaciones de la autoevaluación y evaluación entre iguales, de las rúbricas y de los procesos de realimentación continuada conducidos por profesores, y también por estudiantes. Por otro lado, y por haberse desarrollado en el contexto universitario de las artes, el estudio contribuye a paliar un déficit existente en la literatura especializada: existen pocas investigaciones sobre el campo de las artes en la educación superior, y menos aún conducidas por profesionales de este ámbito de conocimiento.

REFERENCIAS

- ANDRADE, Heidi (2005), "Teaching with Rubrics: The good, the bad, and the ugly", *College Teaching*, vol. 53, núm. 1, pp. 27-31.
- BAUGHIN, Judith, Evelyn Brod y Deborah Page (2002), "Primary Trait Analysis: A tool for classroom-based assessment", *College Teaching*, vol. 50, núm. 2, pp. 75-80.
- BEAUMONT, Chris, Michelle O'Doherty y Lee Shannon (2011), "Reconceptualising Assessment Feedback: A key to improving student learning?", *Studies in Higher Education*, vol. 36, núm. 6, pp. 671-687.
- BOSTOCK, Stephen (2000), "Student Peer Assessment", The Higher Education Academy, en: https://www.cs.auckland.ac.nz/courses/compsc1747s2c/lectures/paul/Student_peer_assessment_-_Stephen_Bostock.pdf(consulta: 8 de febrero de 2012).
- BOUD, David (1990), "Assessment and the Promotion of Academic Values", *Studies in Higher Education*, vol. 15, núm. 1, pp. 101-111.
- BOUD, David (2000), "Sustainable Assessment: Rethinking assessment for the learning society", *Studies in Continuing Education*, vol. 22, núm. 2, pp. 151-167.
- BOUD, David, Ruth Cohen y Jane Sampson (1999), "Peer Learning and Assessment", *Assessment & Evaluation in Higher Education*, vol. 24, núm. 4, pp. 413-426.
- BOUD, David, Ruth Cohen y Jane Sampson (2001), "Peer Learning and Assessment", en David Boud, Ruth Cohen y Jane Sampson (eds.), *Peer Learning in Higher Education: Learning from and with each other*, Londres, Kogan Page, pp. 67-81.
- BOUD, David y Nancy Falchikov (2005), "Redesigning Assessment for Learning beyond Higher Education", ponencia presentada en el "28th HERDSA Annual Conference. Higher Education in a Changing World", Sydney, 3-6 de julio, en: <http://www.hersda.org.au/wp-content/uploads/conference/2005/papers/boud.pdf> (consulta: 23 de junio de 2012).
- BOUZIDI, L'hadi y Alain Jaillet (2009), "Can Online Peer Assessment be Trusted?", *Educational Technology & Society*, vol. 12, núm. 4, pp. 257-268.
- BREW, Angela (1999), "Towards Autonomous Assessment: Using self-assessment and peer assessment", en Sally Brown y Angela Glasner

- (eds.), *Assessment Matters in Higher Education: Choosing and using diverse approaches*, Buckingham, Open University Press, pp. 159-171.
- CHO, Kwangsu, Christian Schunn y Roy Wilson (2006), "Validity and Reliability of Scaffolded Peer Assessment of Writing from Instructor and Student Perspectives", *Journal of Educational Psychology*, vol. 98, núm. 4, pp. 891-901.
- CLARKE, Patricia, Daniel Schull, Glen Coleman, Rachel Pitt y Catherine Manathunga (2013), "Enhancing Professional Writing Skills of Veterinary Technology Students: Linking assessment and clinical practice in a communications course", *Assessment & Evaluation in Higher Education*, vol. 38, núm. 3, pp. 273-287.
- DOCHY, Filip, Mien Segers y Dominique Sluijsmans (1999), "The Use of Self-, Peer- and Co-assessment in Higher Education: A review", *Studies in Higher Education*, vol. 24, núm. 3, pp. 331-350.
- DUNBAR, Norah, Catherine Brooks y Tara Kubicka-Miller (2006), "Oral Communication Skills in Higher Education: Using a performance-based evaluation rubric to assess communication skills", *Innovative Higher Education*, vol. 31, núm. 2, pp. 115-128.
- EL-MOWAFY, Ahmed (2014), "Using Peer Assessment of Fieldwork to Enhance Students' Practical Training", *Assessment & Evaluation in Higher Education*, vol. 39, núm. 2, pp. 223-241.
- FALCHIKOV, Nancy (2007), "The Place of Peers in Learning and Assessment", en David Boud y Nancy Falchikov (eds.), *Rethinking Assessment in Higher Education: Learning for the longer term*, Londres, Routledge, pp. 128-143.
- FALCHIKOV, Nancy y Judy Goldfinch (2000), "Student Peer Assessment in Higher Education: A meta-analysis comparing peer and teacher marks", *Review of Educational Research*, vol. 70, núm. 3, pp. 287-322.
- HANRAHAN, Stephanie y Geoff Isaacs (2001), "Assessing Self- and Peer-Assessment: The students' views", *Higher Education Research & Development*, vol. 20, núm. 1, pp. 53-70.
- HATTIE, John y Helen Timperley (2007), "The Power of Feedback", *Review of Educational Research*, vol. 77, núm. 1, pp. 81-112.
- HORSTMANSHOF, Louise y Sonya Brownie (2013), "A Scaffolded Approach to Discussion Board use for Formative Assessment of Academic Writing Skills", *Assessment & Evaluation in Higher Education*, vol. 38, núm. 1, pp. 61-73.
- JONSSON, Anders y Gunilla Svingby (2007), "The Use of Scoring Rubrics: Reliability, validity and educational consequences", *Educational Research Review*, vol. 2, pp. 130-144.
- KEARNEY, Sean (2013), "Improving Engagement: The use of 'Authentic self-and peer-assessment for learning' to enhance the student learning experience", *Assessment & Evaluation in Higher Education*, vol. 38, núm. 7, pp. 875-891.
- LI, Lan y Allen Steckelberg (2005), "The Instructional Design Portfolio: Peer assessment support system (Pass)", *TechTrends*, vol. 49, núm. 4, pp. 80-84.
- LINDBLOM-YLANNE, Sari, Heikki Pihlajamäki y Toomas Kotkas (2006), "Self-, Peer- and Teacher-Assessment of Student Essays", *Active Learning in Higher Education: The Journal of the Institute for Learning and Teaching*, vol. 7, núm. 1, pp. 51-62.
- MAXWELL, Sherridan (2010), *Using Rubrics to Support Graded Assessment in a Competency Based Environment. Occasional Paper*, Adelaide, National Centre for Vocational Education Research.
- MENÉNDEZ-VARELA, José Luis y Eva Gregori-Giralt (2015a), "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*, vol. 41, núm. 2, pp. 228-244.
- MENÉNDEZ-VARELA, José Luis y Eva Gregori-Giralt (2015b), "Rúbricas para la participación de estudiantes universitarios de reciente ingreso en la evaluación de proyectos de arte", ponencia presentada en el "3rd International Congress of Educational Sciences and Development", San Sebastián, 24-26 de junio de 2015.
- MENÉNDEZ-VARELA, José Luis y Eva Gregori-Giralt (2015c), "La influencia de la evaluación entre iguales basada en rúbricas en el desarrollo de proyectos de arte", ponencia presentada en el "III Congreso Internacional sobre Aprendizaje, Innovación y Competitividad", Madrid, 14-16 de octubre de 2015.
- MONTERO, Ignacio y Orfelio León (2007), "Guía para nombrar los estudios de investigación en Psicología", *International Journal of Clinical and Health Psychology*, vol. 7, núm. 3, pp. 847-862.
- MOSTERT, Markus y Jen Snowball (2013), "Where Angels Fear to Tread: Online peer-assessment in a large first-year class", *Assessment & Evaluation in Higher Education*, vol. 38, núm. 6, pp. 674-686.
- NESTEL, Debra, Roger Kneebone y Jane Kidd (2005), "Triadic Peer Review in Scenario-Based Assessments", *Learning and Teaching in Higher Education*, vol. 1, núm. 1, pp. 108-111.
- NORDRUM, Lene, Katherine Evans y Magnus Gustafsson (2013), "Comparing Student Learning Experiences of In-Text Commentary and Rubric-Articulated Feedback: Strategies for formative assessment", *Assessment & Evaluation in Higher Education*, vol. 38, núm. 8, pp. 919-940.
- O'DONOVAN, Berry, Margaret Price y Chris Rust (2004), "Know what I Mean? Enhancing

- student understanding of assessment standards and criteria”, *Teaching in Higher Education*, vol. 9, núm. 3, pp. 325-335.
- ORSMOND, Paul, Stephen Merry y Kevin Reiling (2000), “The Use of Student Derived Marking Criteria in Peer and Self-Assessment”, *Assessment and Evaluation in Higher Education*, vol. 25, núm. 1, pp. 23-38.
- PARKES, Kelly (2010), “The Use of Criteria Specific Performance Rubrics for Student Self-Assessment: A case study”, en Timothy Brophy (ed.), *The Practice of Assessment in Music Education: Frameworks, models, and designs*, Chicago, GIA Publications, pp. 453-458.
- POPE, Nigel (2001), “An Examination of the Use of Peer Rating for Formative Assessment in the Context of the Theory of Consumption Values”, *Assessment and Evaluation in Higher Education*, vol. 26, núm. 3, pp. 235-246.
- REYNOLDS, Julie, Robin Smith, Cary Moskovitz y Amy Sayle (2009), “BioTAP: A systematic approach to teaching scientific writing and evaluating undergraduate theses”, *BioScience*, vol. 59, núm. 10, pp. 896-903.
- SADLER, D. Royce (2009), “Indeterminacy in the Use of Preset Criteria for Assessment and Grading”, *Assessment & Evaluation in Higher Education*, vol. 34, núm. 2, pp. 159-179.
- SHEPHERD, Lorrie (2000), “The Role of Assessment in Learning Culture”, *Educational Researcher*, vol. 29, núm. 7, pp. 4-14.
- SINGH, Kalayane y J. Terry (2008), “Fostering Students’ Self-Assessment Skills for Sustainable Learning”, ponencia presentada en el “EDUCOM 2008 International Conference. Sustainability in Higher Education: Directions for Change”, Perth, 19-21 de noviembre de 2008, en: <http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1038&context=ceducum> (consulta: 23 de junio de 2012).
- SLUIJSMANS, Dominique, Filip Dochy y George Moerkerke (1999), “Creating a Learning Environment by Using Self-, Peer- and Co-assessment”, *Learning Environments Research*, vol. 1, núm. 3, pp. 293-319.
- SMITH, Calvin Douglas, Kate Worsfold, Lynda Davies, Ron Fisher y Ruth McPhail (2013), “Assessment Literacy and Student Learning: The case for explicitly developing students’ assessment literacy”, *Assessment & Evaluation in Higher Education*, vol. 38, núm. 1, pp. 44-60.
- STUPANS, Ieva, Geoff March y Susanne Owen (2013), “Enhancing Learning in Clinical Placements: Reflective practice, self-assessment, rubrics and scaffolding”, *Assessment & Evaluation in Higher Education*, vol. 38, núm. 5, pp. 507-519.
- THOMAS, Glyn, Dona Martin y Kathleen Pleasants (2011), “Using Self- and Peer-Assessment to Enhance Students’ Future Learning in Higher Education”, *Journal of University Teaching and Learning Practice*, vol. 8, núm. 1, pp. 1-17.
- TOPPING, Keith (1998), “Peer-Assessment between Students in Colleges and Universities”, *Review of Educational Research*, vol. 68, núm. 3, pp. 249-276.
- TOPPING, Keith (2003), “Self- and Peer-Assessment in School and University: Reliability, validity and utility”, en Mien Segers, Filip Dochy y Eduardo Cascallar (eds.), *Optimizing New Modes of Assessment: In search of qualities and standards*, Dordrecht, Kluwer Academic Publishers, pp. 55-87.
- TUCKER, Richard (2013), “The Architecture of Peer Assessment: Do academically successful students make good teammates in design assignments?”, *Assessment & Evaluation in Higher Education*, vol. 38, núm. 1, pp. 74-84.
- VENNING, Jackie y Femke Buisman-Pijlman (2013), “Integrating Assessment Matrices in Feedback Loops to Promote Research Skill Development in Postgraduate Research Projects”, *Assessment & Evaluation in Higher Education*, vol. 38, núm. 5, pp. 567-579.
- YOUNG, Catherine (2013), “Initiating Self-Assessment Strategies in Novice Physiotherapy Students: A method case study”, *Assessment & Evaluation in Higher Education*, vol. 38, núm. 8, pp. 998-1011.