

Este documento está publicado en:

Calzada Prado, F. J. (2020). Avanzar en el aprendizaje autónomo y social: integración de autoevaluación y evaluación por pares como herramientas de evaluación formativa. En *Innovación docente e investigación en educación* (pp. 211-219). Madrid: Dykinson.



El contenido de este documento está sujeto a la licencia [Creative Commons Atribución–Nocomercial–Sinderivadas 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)

CAPÍTULO 18

AVANZAR EN EL APRENDIZAJE AUTÓNOMO Y SOCIAL: INTEGRACIÓN DE AUTOEVALUACIÓN Y EVALUACIÓN POR PARES COMO HERRAMIENTAS DE EVALUACIÓN FORMATIVA

FRANCISCO-JAVIER CALZADA-PRADO
Universidad Carlos III de Madrid

INTRODUCCIÓN

Entre los principales objetivos de la Educación Superior se encuentran el desarrollo de las capacidades de aprendizaje autónomo y social de los estudiantes, por su potencial impacto en el desarrollo integral, tanto profesional, como personal de los individuos. Son muchas las estrategias a disposición del docente para facilitar el desarrollo de estas capacidades en el contexto de la enseñanza de las distintas materias. Habitualmente, las asignaturas contemplan el estudio y realización de actividades formativas orientadas a la adquisición de competencias, tanto de forma individual, como en grupo, en clase -ya sea presencial o virtual- o fuera de ella. Sin embargo, la evaluación de las distintas actividades de aprendizaje queda restringida, en la mayoría de los casos, al ámbito docente: los estudiantes realizan las actividades propuestas según unos presupuestos o criterios concretos y el docente las evalúa, ofrece la retroalimentación que sea pertinente y califica. Esta aproximación a la evaluación puede considerarse una medida eficiente desde el punto de vista de la gestión educativa, pero también puede representar una oportunidad de aprendizaje perdida para los estudiantes.

Diversas investigaciones vienen explorando desde hace algún tiempo (Dochy, Segers, y Sluijsmans, 1999; Falchikov y Boud, 1989; Topping, 1998) la utilidad de la autoevaluación y la evaluación por pares en los procesos de aprendizaje en el ámbito universitario. Los conceptos son ampliamente conocidos, y si bien su práctica en el ámbito universitario no ha sido frecuente, parece estar cobrando importancia desde la adopción de modelos de educación centrados en el estudiante (Wanner y Palmer, 2018). La autoevaluación se ha definido como un proceso mediante el que “se implica a los estudiantes en la identificación de estándares y/o criterios que aplicar a su trabajo y emitir juicios sobre la medida en que han satisfecho tales criterios y estándares” (Boud, 1991; p. 5; citado en Boud, 2005; p. 12). Por su parte, la evaluación por pares como “el conjunto de actividades mediante las cuales los individuos emiten juicios sobre el trabajo de otros” (Reinholz, 2016). Para Black

(2010), la combinación de ambos tipos de evaluación resulta esencial, pues evaluar el trabajo de otros y ver nuestro trabajo a través de los ojos de otros nos permite desarrollar nuestras propias habilidades de autoevaluación. Los beneficios de ambos sobre los procesos de aprendizaje han sido descritos y verificados empíricamente en los últimos años por diversos autores (Logan, 2009; Ndoye, 2017; Wanner y Palmer, 2018), aunque también los condicionantes para su éxito. En este sentido, Wanner y Palmer (2018) señalan, a partir de sus observaciones, que la clave para que estas herramientas de evaluación formativa tengan éxito es un diseño e implementación realizados con la máxima precisión. Para estos autores, “el cambio hacia la autoevaluación y evaluación por pares no es sencillo ni para los estudiantes ni para los profesores, pero merece la pena y es necesario en la educación superior del siglo XXI”.

Objetivo

El objetivo del presente trabajo consiste en presentar los resultados de un proyecto de innovación docente aprobado y reconocido por la Universidad Carlos III de Madrid (15ª Convocatoria de Apoyo a Experiencias de Innovación Docente en Estudios de Grado y Postgrado, Vicerrectorado de Estrategia y Educación Digital), cuyo objetivo principal fue realizar una aproximación a la utilidad de la autoevaluación y la evaluación por pares como instrumentos de evaluación formativa en estudios de Grado y Máster.

MÉTODO

La experiencia se desarrolló durante el curso 2018/2019 sobre dos asignaturas, una de Grado denominada “Information Skills”, del Bachelor’s Degree in Data Science and Engineering, en adelante “IS”, y otra de Máster denominada “Web Social”, del Máster en Bibliotecas, Archivos y Continuidad Digital, en adelante “WS”, contando con un total de 60 estudiantes participantes entre las dos (sobre 68 matriculados). En cada una de las asignaturas se propuso a los estudiantes, como culminación de la evaluación continua, realizar una práctica global individual en la que debían poner en práctica gran parte de las competencias adquiridas a lo largo de la misma. En la asignatura de Grado, a esta práctica se le otorgó un peso del 20% de la evaluación, mientras que en la asignatura de Máster se le asignó un peso del 30%, de acuerdo a las características y requerimientos de cada asignatura.

Para testar los procesos de evaluación previstos en los objetivos del proyecto se empleó la actividad “Taller” (“Workshop”), incorporada ese mismo curso a Aula Global, instancia institucional de Moodle. Taller es una actividad orientada principalmente a facilitar la evaluación por pares entre estudiantes que permite, además, si se desea, incorporar autoevaluación y evaluación docente. El proceso de

evaluación está dirigido por el docente y se desarrolla típicamente en cuatro fases: 1) Configuración del taller, 2) Remisión de tareas y asignación de tareas a evaluar, 3) Evaluación de la entrega, y 4) Calificación y cierre de la actividad.

A estas cuatro fases, prescriptivas en el desarrollo de la actividad Taller, se incorporó en este caso una fase 0 (cero) o preliminar en la que se desarrollaron varias actividades de carácter preparatorio:

1) Una vez explicados los objetivos y procedimiento de realización de la práctica global de asignatura, se proporcionó a los estudiantes recursos y orientación para que comprendieran cómo se iba a desarrollar el proceso de evaluación de la misma, y en especial las implicaciones de la evaluación mediante rúbricas, la evaluación por pares y la autoevaluación, aspectos novedosos para la mayoría de los estudiantes participantes.

2) A continuación, se abrió el proceso de co-creación de una rúbrica para cada una de las prácticas planteadas. Para ello, el profesor utilizó el foro del aula virtual, espacio en el que publicó un primer borrador de la rúbrica con la intención de promover la discusión de los distintos aspectos contemplados en ella (co-responsabilidad). La interacción en el foro entre estudiantes y docente a lo largo de un plazo de una semana generó una versión final de la rúbrica de común acuerdo que sería utilizada por ambas partes para la evaluación de la actividad (evaluación por pares, autoevaluación y evaluación docente).

3) Una vez consolidada la rúbrica, los estudiantes desarrollaron sus prácticas de forma individual y las prepararon para el proceso de evaluación. Para ello, en primer lugar, las subieron al aula virtual a través de una actividad Turnitin, herramienta encargada de realizar una verificación automática de originalidad. En este caso, Turnitin se configuró de manera que los estudiantes pudieran obtener los informes del análisis de originalidad inmediatamente tras la remisión de las prácticas y hasta el fin del plazo de entrega, de manera que pudieran, en caso necesario, mejorar su calidad y garantizar que estuvieran libres de todo tipo de plagio. Por último, los estudiantes anonimizaron sus informes de originalidad Turnitin, eliminando su nombre y cualquier otro elemento que pudiera resultar identificativo. El documento PDF resultante sería posteriormente remitido por los estudiantes a la actividad Taller para iniciar el proceso de evaluación.

Durante la fase 1 o fase de configuración del Taller, se proporcionó una descripción introductoria de carácter general a la actividad y determinaron los tipos de evaluación a realizar e instrumentos de evaluación a aplicar, así como los pesos y calificaciones otorgadas a los distintos elementos. Éstos comprendieron:

1) Una evaluación por pares, en la que cada estudiante debía evaluar dos prácticas de compañeros asignadas aleatoriamente por Taller. La evaluación se configuró como anónima (existe la posibilidad de realizar evaluaciones por pares no

anónimas en Taller), e incluía la evaluación mediante la rúbrica previamente validada, así como dos preguntas de respuesta abierta, con la finalidad de completar el proceso de evaluación de carácter reflexivo: ¿Qué aspectos te han parecido especialmente valiosos de la práctica que has evaluado? ¿Qué propuestas harías a su autor/a para la mejora de la práctica?

2) Una autoevaluación, en la que cada estudiante debía evaluar su propio trabajo utilizando la misma rúbrica y preguntas anteriormente reseñadas.

3) Una evaluación docente a partir de la misma rúbrica y preguntas empleadas por los estudiantes en su evaluación por pares y autoevaluación.

A cada uno de estos tres elementos de evaluación (evaluación por pares, autoevaluación y evaluación docente) se les asignó un peso del 25% de la calificación de la práctica, por lo que el apartado de Taller denominado “Calificación por entrega” tendría un valor agregado del 75%. El 25% de la calificación restante correspondería a la valoración de la calidad de la evaluación por pares (“Calificación por evaluación” o “grading grade”), calculada automáticamente por la actividad. Taller en la actualidad solo permite el método de “Comparación con la mejor evaluación” (“Comparison with best assessment”), que asigna una calificación en función de la proximidad de la calificación otorgada por los estudiantes a los trabajos de sus compañeros a la mejor calificación obtenida por cada trabajo, con distintos grados de sensibilidad, en función de la proximidad estadística: media (“fair”), laxa (“lax”), muy laxa (“very lax”), estricta (“strict”), o muy estricta (“very strict”). En este caso, se optó por una sensibilidad media en el cálculo de calidad de la evaluación.

Una vez finalizada la configuración de la actividad, se inició la segunda fase, en la que los estudiantes pudieron remitir sus prácticas (informes Turnitin anonimizados) y, una vez remitidas, se procedió a realizar la asignación de tareas para la evaluación por pares. En este caso, se optó por una asignación automática aleatoria con la finalidad de que dicha asignación fuera lo más objetiva posible.

En la tercera fase, estudiantes y docente procedieron a realizar sus evaluaciones en un plazo de dos semanas. Se recomendó a los estudiantes comenzar con la evaluación por pares y a continuación, una vez vistos los trabajos de sus compañeros, evaluar su propio trabajo. Completadas las evaluaciones de los estudiantes, el docente procedió a evaluar los trabajos sin tener en cuenta las evaluaciones asignadas previamente por los estudiantes.

Al activar la cuarta y última fase, Taller calcula las calificaciones relativas a calidad de la evaluación por pares. A continuación, el docente revisó las calificaciones asignadas por los estudiantes (en este momento del proceso el docente puede editar calificaciones individuales e incluso variar los pesos de los distintos tipos y elementos de evaluación), así como la calificación por calidad de evaluación

por pares calculada. Una vez finalizado el proceso de revisión, se procedió a cerrar el Taller, momento en el que los estudiantes pudieron acceder a sus calificaciones y retroalimentación facilitadas por estudiantes y docente. Al ser una actividad integrada en la plataforma, las calificaciones del Taller se traspasan automáticamente al libro de calificaciones de Aula Global.

Completado el Taller de evaluación, se trasladó a los estudiantes una encuesta de carácter anónimo para tratar de conocer sus impresiones sobre diversos aspectos relacionados con la utilidad del proceso, obteniendo una tasa media de respuesta en las dos asignaturas de en torno al 80% (76,31% en IS y 86,36% en WS). Las preguntas planteadas a los estudiantes fueron las siguientes:

1. El proceso de evaluación de la práctica de la asignatura me ha ayudado a desarrollar mi capacidad de reflexionar sobre mi trabajo y mejorar en el futuro.

a) Sí; b) No.

2. El proceso de evaluación de la práctica de la asignatura me ha ayudado a desarrollar mi capacidad de reflexionar sobre el trabajo de otros y darles retroalimentación constructiva.

a) Sí; b) No.

3. Considero que la calificación que he obtenido en la práctica ha sido adecuadamente justificada.

a) Sí; b) No.

4. El proceso de evaluación de la práctica me ha ayudado a entender mejor el propósito de la actividad y alcanzar sus objetivos de aprendizaje.

a) Sí; b) No.

5. Saber que mi práctica sería evaluada por mis compañeros me animó a hacer un mejor trabajo.

a) Sí; b) No.

6. Considero que la calificación otorgada por los estudiantes en la evaluación por pares debería tener más peso que la dada por el profesor.

a) Sí; b) No.

7. He entendido cómo se ha calculado la calificación otorgada por la calidad de la evaluación por pares realizada (“grading grade”).

a) Sí; b) No.

8. Saber que una parte de la calificación de la práctica sería calculada teniendo en cuenta la calidad de la evaluación que otorgara contribuyó a esforzarme por realizar una evaluación más precisa del trabajo de mis compañeros.

a) Sí; b) No.

9. La rúbrica me ayudó a realizar la evaluación de la actividad.

a) Sí; b) No.

10. No participé en el desarrollo de la rúbrica porque...

a) Me pareció bien la propuesta del profesor.

b) No sabía cómo hacerlo.

c) No me sentí preparado/a para hacerlo.

11. El uso de la rúbrica me permitió realizar una evaluación justa del trabajo de mis compañeros y el mío propio.

a) Sí; b) No.

12. Las preguntas que seguían a la rúbrica en el proceso de evaluación (centradas en valorar aspectos positivos y aspectos a mejorar) eran relevantes y útiles.

a) Sí; b) No.

13. La información proporcionada para la realización de la autoevaluación y la evaluación por pares fue suficiente para poder hacer la actividad.

a) Sí; b) No.

El cuestionario se configuró de manera que no fuera obligatorio contestar a todas las preguntas, por lo que los estudiantes podían decidir qué preguntas contestar o no, si bien la mayoría de los participantes respondieron a casi todas las preguntas.

RESULTADOS

El proceso de evaluación de las prácticas con Taller se desarrolló a lo largo de las etapas previstas en el tiempo establecido, por lo que se estima que los tiempos considerados para su realización (unas seis semanas en total) resultaron adecuados. Desde el punto de vista docente, resultó especialmente interesante observar que la evaluación por pares fue completada por los estudiantes de ambas asignaturas en los primeros días de la fase de evaluación, lo cual podría reflejar su especial interés en este tipo de evaluación. En el caso de la autoevaluación se observó justo lo contrario, los estudiantes tendieron a dejarla para el final de la fase de evaluación, lo que sugiere, junto con algunos comentarios en las autoevaluaciones, una mayor dificultad en su realización.

Atendiendo a las calificaciones otorgadas (Tabla 1), observamos varias cuestiones:

1) Las calificaciones asignadas por los estudiantes a sus propios trabajos en el proceso de autoevaluación fueron más altas que las otorgadas por sus compañeros en la evaluación por pares (2,8% tanto en IS como en WS) y por el docente en su evaluación (2,9% en IS y 6,2% en WS). Habría sido interesante ver si los estudiantes calificarían de la misma manera sus trabajos después de tener acceso al conjunto de calificaciones y retroalimentación obtenidas. En el futuro, cabría explorar la posibilidad de ampliar el proceso de evaluación en Taller con una nueva fase de

revisión de la autoevaluación, orientada a completar aún más el proceso de aprendizaje reflexivo.

Tabla 1. Medias de calificaciones otorgadas en los elementos de evaluación considerados

	Evaluación por pares (25%)	Autoevaluación (25%)	Evaluación docente (25%)	Calificación por entrega (75%)	Calificación por evaluación (25%)	Final (100%)
<i>Information Skills</i>	21.1	23,9	21	66	23,7	89,7
Web Social	19.7	22.5	16.3	58,5	23,2	81,7

2) Las calificaciones medias emanadas de la evaluación por pares resultaron casi idénticas a las correspondientes a la evaluación docente en el caso de IS. En el caso de WS, las calificaciones de la evaluación por pares fueron algo más altas que las de la evaluación docente, si bien no tanto como las procedentes de la autoevaluación. Esto parece sugerir una mayor calidad (o más concretamente, precisión) de la evaluación por pares frente a la autoevaluación. Los resultados de la calificación automática generada por taller en función de la calidad estadística de las calificaciones otorgadas por los estudiantes durante la evaluación por pares, por su parte, también sugieren que éstas fueron en general bastante precisas (23,7/25 en IS y 23,2/25 en WS).

Por otra parte, y atendiendo a los resultados de la encuesta realizada, se observó, además, que:

1) Casi todos los estudiantes encuestados de ambas asignaturas (97,4%) estimaron que el proceso de evaluación desarrollado en la actividad les ayudaba a desarrollar su capacidad de reflexionar sobre su trabajo y mejorar en el futuro. Una gran mayoría (94%) consideró que le había ayudado también a desarrollar su capacidad de reflexionar sobre el trabajo de otros y darles retroalimentación para la mejora de su trabajo. Al 81,6% de ellos, el proceso de evaluación incluso les ayudó a entender mejor el propósito de la actividad y alcanzar sus objetivos de aprendizaje. En sus autoevaluaciones, algunos estudiantes de WS comentaron: “Después de haber leído los trabajos de otras personas, he podido comprobar algunas deficiencias en el mío”, “Me ha servido para darme cuenta de que hay maneras muy distintas de enfocar una misma práctica y ver el excelente nivel que tienen mis compañeros”, “Me ha parecido un proceso de aprendizaje valioso”.

2) El uso de la rúbrica como instrumento de evaluación fue especialmente bien valorado por los estudiantes (80,8%), así como las preguntas que las acompañaban (86%), todo lo cual contribuyó a que vieran bien justificada las calificaciones

obtenidas (73,8%). Muchos estudiantes decidieron no participar en la elaboración de la rúbrica porque les pareció bien la propuesta inicial del docente (47%), si bien otros no lo hicieron porque no sabían muy bien cómo hacerlo (13%) o no se sintieron preparados para ello (17,3). Esto puede estar relacionado con otro de los aspectos de la encuesta, ya que si bien una mayoría (73%) afirma que la información proporcionada para la realización de la autoevaluación y la evaluación por pares había sido suficiente para poder hacer la actividad, otro porcentaje (14%) consideraba que no había sido así. Esto parece sugerir que sería necesario asegurar la formación previa en cuanto a diseño de rúbricas.

3) Otra cuestión relevante es el hecho de que la mayoría de los estudiantes (77,4% frente a un 1,7%) consideran que la calificación otorgada por el profesor en la evaluación por pares debería tener más peso que la dada por los estudiantes. Esto puede entenderse como un reflejo de la tradicional figura del docente como 'autoritas' en las materias que imparte. En el futuro, se deberá tener en cuenta este aspecto de manera que la calificación del docente pese más en la calificación en este tipo de evaluación.

4) Saber que iban a ser evaluados por sus compañeros no fue un factor que influyera de forma importante en su deseo de hacer un trabajo de mejor calidad (53%), sino que parece más bien responder a motivaciones intrínsecas. Saber que una parte de su calificación vendría dada por la calidad de la evaluación que ofrecieran a sus compañeros sí ayudó a la mayoría de ellos (87,8%) a esforzarse por realizar una evaluación más precisa, a pesar de que algunos no llegaron a entender del todo cómo Taller calcula dicha calificación (20,5% frente al 67% que sí lo entendió). La calificación por evaluación, pues, parece un elemento importante y a conservar en futuras aplicaciones, si bien se debe garantizar que todos los estudiantes comprenden el método y cálculos aplicados por Taller.

DISCUSIÓN/CONCLUSIONES

A la vista de los resultados de la experiencia desarrollada, podemos afirmar que la integración de evaluación por pares y autoevaluación en la evaluación formativa parece mostrar una clara utilidad en el avance hacia el aprendizaje social y autónomo en el ámbito de la Educación Superior. Los estudiantes han recibido muy positivamente la aplicación de este tipo de procesos de evaluación como parte de su experiencia de aprendizaje, valorando especialmente su contribución a la mejora de su capacidad de aprendizaje y mejora continua o su capacidad de trabajar y aprender con otros. Consideramos, además, que la experiencia ha sido especialmente valiosa por su potencial de amplia aplicación a otras materias, que habrá de ser explorada en mayor detalle en futuras experiencias e investigaciones.

REFERENCIAS

Black, P. (2010). *Formative Assessment*. En P. Peterson, E. Baker y B. McGaw (Eds.), *International Encyclopedia of Education*, (pp. 359-364). California: Elsevier.

Boud, D. (1991). *HERDSA Green Guide No 5. Implementing student self assessment*. Campbelltown: The Higher Education Research and Development Society Australasia (HERDSA).

Boud, D. (2005). *Enhancing Learning Through Self-assessment*. London: Kogan Page.

Dochy, F., Segers, M., y Sluijsmans, D. (1999). The Use of Self-, Peer and Co-assessment in Higher Education: A review. *Studies in Higher Education*, 24(3), 331-350.

Falchikov, N., y Boud, D. (1989). Student Self-Assessment in Higher Education: A Meta-Analysis. *Review of Educational Research*, 59(4), 395-430.

Logan, E. (2009). Self and Peer Assessment in Action. *Practitioner Research in Higher Education*, 3(1), 35.

Ndoye, A. (2017). Peer / Self Assessment and Student Learning. *International Journal of Teaching*, 29(2), 255-269.

Reinholz, D. (2016). The assessment cycle: a model for learning through peer assessment. *Assessment and Evaluation in Higher Education*, 41(2), 301-315.

Topping, K. (1998). Peer Assessment Between Students in Colleges and Universities. *Review of Educational Research*, 68(3), 249-276.

Wanner, T., y Palmer, E. (2018). Formative self-and peer assessment for improved student learning: the crucial factors of design, teacher participation and feedback. *Assessment and Evaluation in Higher Education*, 43(7), 1032-1047.