



RELIEVE. Revista Electrónica de Investigación y
Evaluación Educativa

ISSN: 1134-4032

relievejournal@gmail.com

Universidad de Granada
España

Gómez-Ruiz, M.A.; Quesada-Serra, V.
Análisis de las calificaciones compartidas en la modalidad
participativa de la evaluación colaborativa entre docente y estudiantes

RELIEVE. Revista Electrónica de Investigación y
Evaluación Educativa, vol. 26, núm. 1, 2020, Enero-Junio
Universidad de Granada
Valencia, España

DOI: <https://doi.org/10.7203/relieve.26.1.16567>

Disponible en: <https://www.redalyc.org/articulo.oa?id=91664838007>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica Redalyc

Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso
abierto

Análisis de las calificaciones compartidas en la modalidad participativa de la evaluación colaborativa entre docente y estudiantes

An Analysis of Shared Grading in Co-Assessment Practices by Teachers and Students

Gómez-Ruiz, M.A.⁽¹⁾ , & Quesada-Serra, V.⁽²⁾ 

(1) Universidad de Cádiz (España). (2) Universitat de les Illes Balears (España).

Abstract

Current trends in educational assessment in different branches of higher education share the common goal of uniting learning with assessment. Most approaches and theoretical and practical developments in this field revolve around four main factors: feedback, democratization, alignment and relevance. This paper proposes the use of co-assessment as a means of ensuring dialogue-based, democratic and fairer evaluations. With co-assessment, the responsibility is shared by the teacher and the students, who negotiate and agree on the appraisal of student tasks and, in this paper, also on the awarded mark. The aim of this study is to analyse the relationship between a series of jointly agreed marks, following the co-assessment of four tasks, and the marks that the teachers and students would each have individually awarded. Two teachers and 100 students participated in the study, which follows a correlational design and analyses significant statistical differences. The results show a strong correlation between the jointly agreed marks and those assigned individually by the teacher, even though statistically significant differences were found between them. Conversely, no statistically significant differences were identified between the joint marks and the marks assigned individually by the students. These results call for reflection on the real possibility of adapting shared grading methods to students in university frameworks, where the repercussions of awarded marks go far beyond formative goals.

Keywords: Grading; Student Evaluation; Teacher Student Relationship; Alternative Assessment; Educational assessment; co-assessment

Resumen

Dentro de la variedad existente en las tendencias actuales sobre la evaluación de estudiantes, se encuentra el propósito común de relacionar evaluación y aprendizaje. La retroalimentación, la democratización, la coherencia y la relevancia son cuatro tópicos aglutinadores sobre los que giran la mayoría de los planteamientos y avances teóricos y prácticos en este ámbito. Como forma concreta de cristalizar una evaluación dialógica, democrática y justa, se propone la modalidad participativa de la evaluación colaborativa en la que docentes y estudiantes se reparten la responsabilidad, negociando y consensuando de forma conjunta el valor de las tareas y en nuestro caso, también la calificación final. El propósito de este estudio, que ha involucrado a un total de 100 alumnos y 2 docentes, es precisamente la comprobación del grado de relación existente entre las calificaciones compartidas de 4 tareas universitarias con las que habrían aportado en solitario el docente y el grupo de estudiantes. Se ha seguido un diseño de investigación correlacional y se ha comprobado la existencia de diferencias significativas. Los resultados muestran la estrecha correlación entre las calificaciones compartidas y las calificaciones del docente, aunque se han hallado diferencias estadísticamente significativas entre estas. Por otro lado, no se han encontrado diferencias entre las calificaciones compartidas y las calificaciones de los estudiantes. Las repercusiones de estos resultados, hacen reflexionar, entre otras cuestiones, sobre la posibilidad real de ajustar dichas calificaciones abiertas a la participación de los estudiantes en contextos universitarios donde las repercusiones sobrepasan de largo el ámbito únicamente formativo.

Palabras clave: Calificación; Evaluación de estudiantes; Relación de docentes y estudiantes; Evaluación alternativa; Evaluación educativa; Coevaluación

Received/Recibido 2020 January 27 | Approved/Aprobado 2020 June 03 | Published/Publicado 2020 June 30

Corresponding author / Autor de contacto: Gómez-Ruiz, M.A. Universidad de Cádiz, Avda. República Saharaui s/n, Campus de Puerto Real; 11519, Puerto Real, Cádiz, España). miguel.gomez@uca.es

En las últimas décadas, los avances científicos teórico-prácticos en el ámbito de la evaluación de aprendizajes, mayormente enfocados en la Educación Superior, han traído consigo una variedad de conceptualizaciones y nuevos enfoques. Desde la ya, quizá tradicional, evaluación formativa, se ha ido diversificando el énfasis en uno u otro aspecto para crear múltiples corrientes y teorías, que van desde la evaluación orientada al aprendizaje (Carless, 2007), hasta la evaluación sostenible (Boud & Soler, 2015), la evaluación como aprendizaje y empoderamiento (Rodríguez & Ibarra, 2015), la evaluación integrada (Crisp, 2012), la evaluación formativa y compartida (López, 2009; 2012) o incluso, desde perspectivas más sociales y críticas, la evaluación inclusiva (Santiuste & Arranz, 2009) o la evaluación para la justicia social (Hidalgo & Murillo, 2016), solo por citar algunos ejemplos.

Probablemente sean muchos los adjetivos y excesivas tanto las particularidades como las reiteraciones, pero dentro de esta variedad, en la base de la mayoría de estas teorizaciones, se encuentra un propósito común de concebir la evaluación como una realidad estrechamente ligada al aprendizaje de los implicados -tanto estudiantes como docentes-, ya sea para confirmarlo, adaptarlo o desarrollarlo (Hayward, 2015; Ibarra & Rodríguez, 2019). Igualmente, las estrategias e ideas fuerza de las actuales concepciones de la evaluación comparten distintos nexos en sus planteamientos que podríamos resumir, de manera aglutinadora, en cuatro tópicos: la retroalimentación, la democratización, la coherencia y la relevancia (Figura 1).

La retroalimentación. En todos los enfoques se coincide en que para desarrollar una evaluación formativa es necesaria la existencia de una retroalimentación o *feedback* que provea a los estudiantes de una información de calidad y adecuada en tiempo y forma sobre sus desempeños, considerándose, de hecho, como el proceso más relevante en la enseñanza para fomentar el aprendizaje y desarrollar la autorregulación de los estudiantes (Hattie & Timperley, 2007). Además, en los últimos años se ha destacado la concepción de la retroalimentación como actividad generalizable

a otras tareas o ámbitos, una retroalimentación transferible o *feedforward* que enfatiza una evaluación con sentido prospectivo y constructivo (Canabal & Margalef, 2017).

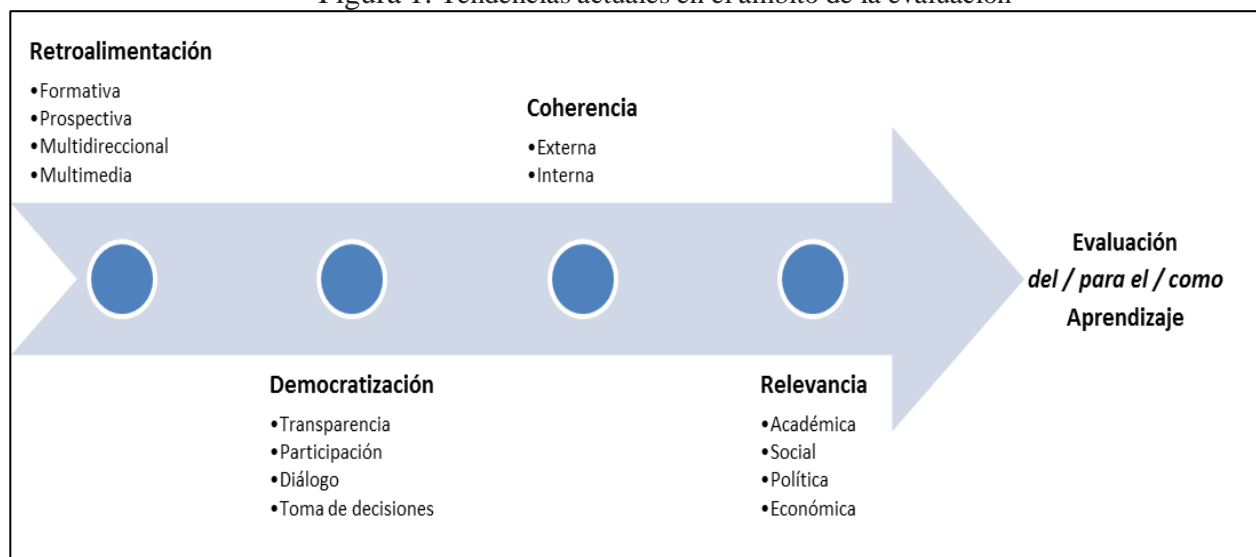
No obstante, también ha existido una evolución del concepto de retroalimentación respecto a los agentes participantes y los medios en los que se aporta. Respecto a los agentes, de una visión unidireccional de la información desde el docente como experto hacia el estudiante como aprendiz, se ha pasado a considerarlo un proceso multidireccional donde los estudiantes cobran protagonismo, ya sea para aportar información útil a sus compañeros o al propio docente (López & Sicilia, 2017; Nicol, 2010). Los estudiantes pueden retroalimentar al docente (Swaffield, 2011) para que le permita considerar y reflexionar sobre la metodología de enseñanza utilizada, la relevancia de las tareas de evaluación, la consecución de objetivos académicos o las repercusiones de la propia retroalimentación que emite a sus estudiantes. Igualmente, propiciar que los estudiantes se aporten retroalimentación entre sí es una práctica que no solo puede beneficiar al receptor, al considerarla una información más comprensible, cercana y útil que la proporcionada por el docente (Gallego et al., 2017; Topping, 2003), sino que en los emisores favorece el desarrollo de capacidades argumentativas y evaluativas (Nicol, 2013). Respecto a los medios de soporte, las tecnologías han permitido pasar de la retroalimentación dialógica síncrona y la escrita asíncrona, a una diversidad de formatos y combinaciones multimedia de estos, en los que destacan por su potencial las opciones de audio y vídeo (García et al., 2015).

La democratización. Se acepta de manera explícita que evaluar y calificar son, en sí mismos y en la educación formal, actos de poder (Leach et al., 2010). Hace tiempo que se considera que, de una u otra manera, ese poder institucional debe redistribuirse, al menos parcialmente, entre los estudiantes (Quesada et al., 2019). Desde un nivel informativo, se habla de la necesidad de que la evaluación sea transparente (Ibarra & Rodríguez, 2019) haciendo público, accesible y comprensible los sistemas de evaluación al completo. Pero

también desde este foco de interés, aparece con insistencia la importancia de la implicación y participación del alumnado en los procesos de evaluación, ya sea en el diseño (criterios y planificación), la ejecución (como responsables

de la evaluación mediante las modalidades participativas de la autoevaluación, evaluación entre iguales o evaluación colaborativa) o la decisión sobre la calificación, de manera única o consensuada (Quesada et al., 2017).

Figura 1. Tendencias actuales en el ámbito de la evaluación



Nota: Elaboración propia.

Sea de una u otra forma, la participación de los estudiantes en la evaluación pretende que estos perciban que tienen un papel fundamental en la construcción de su propio aprendizaje, fomentando su autonomía y aumentando su responsabilidad (Penuel & Shepard, 2016). Igualmente, el diálogo, el consenso y la toma conjunta de decisiones en el diseño y el desarrollo de los procesos de evaluación de los implicados son aspectos recurrentes en las distintas teorizaciones actuales (López & Sicilia, 2017), estableciéndose, igualmente, la pertinencia de una alfabetización del alumnado en la evaluación como vía para garantizar una implicación razonada, consecuente y eficaz de estos (Smith et al., 2013).

La coherencia. Tradicionalmente, desde la perspectiva del docente, la evaluación era el último elemento a considerar en los diseños didácticos tras la especificación de los objetivos y las actividades de enseñanza. No obstante, en la actualidad la evaluación suele abordarse centrada desde el prisma de los estudiantes, para los cuales supone su punto de partida para realizar las actividades de aprendizaje a la espera de resultados positivos (Biggs & Tang,

2011). Se plantea, por tanto, dar coherencia a todo el proceso (coherencia externa), alineando de forma consecuente los objetivos, la metodología, las actividades y la evaluación planteada; prestándose especial atención al papel que juegan las tareas de aprendizaje y evaluación en el proceso, ya que definen el quehacer de los estudiantes, por lo que se proponen que sean auténticas y realistas (Swaffield, 2011), es decir, aprovechables y útiles para su formación.

Por otro lado, también se insiste en la importancia de definir y alinear los distintos componentes de la evaluación (coherencia interna): criterios, medios, productos, técnicas e instrumentos, conectándolos congruentemente y sistematizando lo máximo posible los procedimientos de evaluación (Ibarra & Rodríguez, 2019; Quesada et al., 2017).

La relevancia. La evaluación de aprendizajes se ha convertido en un ámbito de investigación cada vez más profundo y variado. Las evidentes consecuencias académicas y cognitivas de las prácticas evaluativas hace tiempo que le dieron

una relevancia primordial en los procesos de enseñanza y aprendizaje en todos los niveles educativos. No obstante, también se están considerando cada vez más frecuentemente hasta qué punto la evaluación afecta a la vida de los estudiantes (McArthur, 2019) y cuáles son sus verdaderas repercusiones sociales, emocionales, políticas y económicas. Esta relevancia pluridimensional, la conecta directamente con el marco axiológico del docente y le hace reflexionar sobre las consecuencias para el presente y el futuro, no solo de nuestro alumnado, sino de la sociedad que estamos construyendo (Hidalgo & Murillo, 2016).

Evaluación colaborativa como modalidad participativa

Sería imposible establecer una única cristalización práctica del actual desarrollo teórico de la evaluación, por la gran variedad de combinaciones posibles, por el énfasis en uno u otro aspecto o por la propia naturaleza limitada e imperfecta de las experiencias evaluativas en la realidad. No obstante, consideramos que la evaluación colaborativa, transparente en su diseño y procesual, formativa y orientadora en su implementación, puede ser una alternativa factible para conseguir una evaluación democrática que fomente la participación y la toma de decisiones, donde la retroalimentación dialogada se coloca en el centro del proceso, se alinea de manera coherente con la metodología y las tareas de evaluación, y donde se asume la orientación inclusiva, crítica, optimista y justa que permita convertir la evaluación en una estrategia orientada hacia la justicia social (Murillo & Hidalgo, 2015).

El concepto de evaluación colaborativa es también denominado en la literatura científica como coevaluación, co-evaluación o evaluación compartida. No obstante, los términos coevaluación y co-evaluación, traducción literal de *co-assessment*, son también utilizados en español para referirse a la evaluación entre iguales, mientras que evaluación compartida puede igualmente hacer referencia a otros marcos más amplios (véase el caso de López, 2012).

Por ello, y con la finalidad de facilitar la claridad en las publicaciones científicas, en este trabajo proponemos la utilización del término evaluación colaborativa para referirnos a aquella modalidad participativa de evaluación que fomenta el consenso a través de una apreciación y discusión conjunta entre docentes y estudiantes de la tarea del estudiante. Las otras dos modalidades de evaluación participativa existentes son la autoevaluación y la evaluación entre iguales, aunque es la evaluación colaborativa la más desconocida, la menos utilizada y la que más inseguridad produce entre el profesorado universitario (Quesada et al., 2016).

Así, se entiende la evaluación colaborativa como aquella evaluación que es negociada conjuntamente por docentes y estudiantes con el objetivo de alcanzar un consenso, siendo la responsabilidad compartida y el diálogo esencial. De esta definición son clave tres conceptos: responsabilidad, diálogo y consenso, que reflejan una manera de entender las interacciones educativas desde un prisma de compromiso político y social. El docente aquí no es el único poseedor de la verdad y se considera que el alumnado es capaz de calibrar sus aprendizajes, tomar decisiones y argumentarlas, cambiando así la identidad de los estudiantes, concebidos como personas que construyen aprendizajes y responsables de su situación académica. (Quesada et al., 2019).

La evaluación colaborativa pues, puede considerarse que está compuesta por una parte de autoevaluación del estudiante al evaluar su propio trabajo, evaluación entre iguales, cuando además se trabaja en grupo, y otra parte de evaluación del profesorado que se combinan dialógicamente de forma explícita (Quesada et al., 2016). Siguiendo esta línea, Kurt (2014) entiende la evaluación colaborativa directamente como una combinación de autoevaluación, evaluación entre iguales, evaluación del profesorado y una evaluación negociada. No obstante, es necesario no confundir la evaluación colaborativa con aquellas prácticas en las que el estudiante realiza una autoevaluación y los docentes evalúan la tarea de manera separada, sin que

tenga lugar un proceso de diálogo y negociación. Es precisamente ese proceso de confrontación o triangulación el que permite al profesorado obtener información relevante sobre cómo se está desarrollando el proceso de aprendizaje, y al alumnado, mayor posibilidad de aprendizaje y autorregulación, al ofrecerse la oportunidad de formarse una idea más concreta de su propia actuación y de las valoraciones ofrecidas.

Algunos de los beneficios en el aprendizaje del alumnado que se le atribuyen en la literatura especializada a la práctica de la evaluación colaborativa son (Boud & Falchikov, 2006; Cooper, 2017; Deeley, 2014; Dochy et al., 1999; Gómez & Quesada, 2017; Knight & Yorke, 2003; Quesada et al., 2016; Quesada et al., 2019): Promueve la reflexión sobre el propio desempeño y un aprendizaje más profundo; desarrolla la autorregulación, fomenta la autonomía y la toma de decisiones; aumenta la relación y comunicación entre docente y alumnado; acrecienta el esfuerzo, la motivación y la implicación de los participantes; mejora la alfabetización evaluadora, así como la autoestima y la confianza en la propia capacidad para llevar a cabo la evaluación.

No obstante, el uso de esta modalidad evaluativa puede conllevar también algunos riesgos e inconvenientes (Ibídem), tanto para el profesorado como para el alumnado participante. Por parte del alumnado, se destaca la complejidad que implica evaluar el propio trabajo, así como la incomodidad y tensión que supone para algunos estudiantes el diálogo sobre su desempeño con el docente. Para el profesorado puede acarrear una mayor carga de trabajo, incluso con ratios reducidas, a la vez que se reconoce como una práctica de complicada implementación en grupos numerosos.

La calificación compartida de las tareas evaluadas

La práctica de la evaluación colaborativa puede, o no, incluir el consenso de una calificación conjunta y negociada. No es un requisito imprescindible, pues los aspectos más

relevantes son el diálogo, la argumentación y las evidencias que fundamenten las valoraciones de las actuaciones y la toma de decisiones conjuntas (Van der Bergh et al., 2006). La culminación de los procesos colaborativos de evaluación en una calificación compartida es una posibilidad que propicia muchos debates y levanta ciertas suspicacias entre los investigadores, suponiéndose una desnaturalización del proceso que puede llegar a desvirtuarse debido a la sobrevaloración o infravaloración de las calificaciones compartidas finales. También existen autores que sostienen realizar experiencias de calificación dialogada únicamente al final de las asignaturas (López, 2012).

Plantearse explícitamente el papel de la calificación en las nuevas conceptualizaciones de la evaluación es una realidad incómoda para muchos docentes e investigadores en este ámbito. En ocasiones parece obviarse de los enfoques teóricos formativos de la evaluación, ya sea intencionada o inintencionadamente. Igualmente, se puede sobreentender que sigue siendo una responsabilidad exclusiva del docente o incluso que puede ser perjudicial que entre en juego el alumnado, pues puede corromper el proceso de aprendizaje. No obstante, al menos en el ámbito universitario, calificar el desempeño de los estudiantes en las asignaturas con un único número es una exigencia institucional y legal. Los docentes seguimos teniendo la obligación de traducir los procesos y productos educativos de cada uno de nuestros estudiantes en una calificación y debemos decidir si incluimos al alumnado, como parte de la participación activa, en la toma de decisiones de sus propias calificaciones y las de sus compañeros.

Por un lado, integrar a los estudiantes en las calificaciones puede significar una involucración más profunda, auténtica y significativa en el proceso de enseñanza, aprendizaje y evaluación. Es decir, se puede considerar un signo inequívoco de la implicación y decisión sobre su propio aprendizaje (Álvarez, 2001). También se pueden conseguir mayores cotas de democratización, pues los estudiantes

participan en todas las decisiones que le afectan directamente y podrían apreciar cierta lógica y continuidad de la evaluación y la calificación para utilizarse de manera racional y consecuente, alejándola de las realidades simplistas que fomentan la competitividad, la jerarquización, el etiquetado e incluso la marginación del alumnado (Casanova, 2011).

No obstante, no se puede obviar que algunas experiencias de investigación en el campo han puesto de manifiesto que esta integración en las calificaciones puede provocar que ciertos estudiantes intenten buscar estrategias que les permita incrementar su calificación durante el proceso de negociación (Véase Deeley, 2014), de hecho, incluso en trabajos donde se exploran las opiniones de los propios estudiantes participantes en los procesos de evaluación colaborativa (Gómez & Quesada, 2017; Quesada et al., 2019) estos también manifiestan preocupación sobre la posibilidad de que exista una falta de ajuste en la calificación por su participación en la toma de decisiones. Preocupaciones, por el momento, a las que no se ha dado respuesta adecuada ni definitiva a través de la evidencia científica.

En la literatura ha sido habitual encontrar estudios que analizan la fiabilidad y validez de las calificaciones aportadas en las modalidades de la autoevaluación y evaluación entre iguales (Bretones, 2008; Falchikov & Goldfinch, 2000). No obstante, la utilización minoritaria de la evaluación colaborativa y la calificación compartida, así como su propia naturaleza, nos impiden encontrar investigaciones igualmente sólidas sobre la posible relación entre las calificaciones compartidas y las que aportarían estudiantes y profesorado por separado.

Objetivos

Precisamente en este trabajo presentamos un estudio que tiene como finalidad central valorar las posibles divergencias entre las calificaciones compartidas durante la evaluación colaborativa y las calificaciones que se habrían otorgado los estudiantes que realizan la tarea de evaluación o el docente en solitario.

De forma concreta, se enumeran a continuación los objetivos de esta investigación:

- Analizar el grado de correlación existente entre las calificaciones compartidas derivadas de una evaluación colaborativa y dialógica, con las calificaciones que habría aportado el docente y el grupo de estudiantes que realizó la tarea, de manera independiente.
- Comprobar si existen diferencias estadísticamente significativas entre la calificación compartida y las calificaciones que hubieran aportado el docente o los estudiantes a las tareas realizadas.

Método

El diseño de investigación utilizado en este estudio es de carácter no experimental, optando por un diseño correlacional, en el que los investigadores no han podido controlar ni manipular las variables de estudio, intentando establecer la fuerza y dirección de la relación entre las distintas calificaciones analizadas (Hernández & Maquilón, 2010). Este diseño se complementa con un estudio de diferencia de medias.

Procedimiento

Teniendo en cuenta el enfoque práctico de la investigación, se considera especialmente relevante detallar cómo fue el proceso completo de esta experiencia evaluativa. En ella, los docentes responsables propusieron evaluar y calificar conjuntamente las 4 tareas grupales -2 cada uno- que suponían un 30% de la calificación total de la asignatura en la que se desarrolló. Estas prácticas fueron las siguientes:

- Tarea 1: Respuesta creativa a cuestiones mediante audio, imágenes o vídeo después del análisis de varios recursos escritos y audiovisuales.
- Tarea 2: Realización de un ensayo sobre el tópico de la importancia de la reflexión en la innovación docente.
- Tarea 3: Diseño de investigación parte I. Elección del tema de investigación,

redacción de un breve marco teórico y formulación de preguntas de investigación.

- Tarea 4: Diseño de investigación parte II. Metodología: muestra, temporalización, técnicas e instrumentos de recogida de información.

El proceso seguido para la evaluación y calificación de cada tarea ha sido adaptado de otras experiencias similares (Gómez & Quesada, 2017; Quesada et al., 2019), aunque introduciendo modificaciones inevitables para recoger las distintas calificaciones con el fin de conseguir los objetivos de esta investigación. A continuación, enumeramos los pasos o etapas abordados:

i. Presentación de la tarea en clase. Lo primero que hacía cada docente era especificar el contenido de la actividad y detallar claramente los criterios con los que se evaluaría conjuntamente la calidad de dicha tarea.

ii. Realización de la práctica. Los estudiantes disponían de dos o tres semanas para su ejecución y entrega mediante el Campus Virtual de la asignatura.

iii. Revisión y calificación del docente. Tras la entrega, el docente las corregía y anotaba comentarios en el propio documento que servirían de base para la retroalimentación. Posteriormente asignaba una calificación a la tarea y la anotaba en la ficha de registro.

iv. Reunión de coevaluación. Durante la siguiente sesión práctica presencial, el docente se reunía con cada grupo de trabajo y dedicaba aproximadamente entre 15 y 20 minutos a evaluar colaborativamente y de forma dialogada y razonada la práctica entregada mediante las siguientes fases:

- a. Lo primero era que los estudiantes aportaran sus ideas y opiniones previas respecto a la tarea realizada.
- b. Posteriormente, se mostraban y explicaban los comentarios realizados por el docente, atendiendo a los criterios de evaluación, en forma de retroalimentación.
- c. Posteriormente se abrían las intervenciones de los estudiantes en las que mostraban su acuerdo o desacuerdo con las apreciaciones del docente, se resolvían las dudas, exponían sus percepciones y opiniones sobre las dificultades encontradas en el proceso y sobre la pertinencia del producto entregado.
- d. Tras este diálogo entre estudiantes y docente, el profesor invitaba a los estudiantes a reflexionar sobre la calificación que asignarían a la tarea. En este momento los componentes del grupo dejaban constancia de su calificación de manera individual en un papel.
- e. Por último, se abría un nuevo momento de diálogo, intercambio y negociación entre los estudiantes y entre estos con el docente para consensuar, a partir de sus apreciaciones personales, la calificación definitiva de forma compartida. Al concluir la reunión, el docente se encargaba de anotar en la ficha de registro las calificaciones individuales de cada componente del grupo y la calificación final de la tarea.

De forma ilustrativa, ofrecemos en la Figura 2 un resumen del proceso:

Figura 2. Proceso de la experiencia de calificación compartida



Participantes

La recogida de información se desarrolló en la asignatura “Innovación e Investigación Educativas”, una materia obligatoria situada en el primer semestre del segundo curso del Grado en Educación Primaria impartido en la Universidad de Cádiz durante el año académico 2016-2017. La selección de los participantes en este estudio se desarrolló mediante un muestreo no probabilístico e intencional (Wood & Smith, 2018) basados en criterios de oportunidad de acceso para los investigadores.

Concretamente, participaron un total de 100 estudiantes divididos en 3 subgrupos de prácticas, de los cuales 29 eran hombres (29%)

y 71 mujeres (71%). Los estudiantes se agrupaban en 22 equipos de trabajo, siendo lo más común las agrupaciones de 4 a 6 personas (86,4%). Es relevante señalar que estos estudiantes ya habían cursado anteriormente otra asignatura con uno de los docentes - Docente A- en la que también tuvieron experiencias relacionadas con la evaluación colaborativa (incluyendo calificación compartida), por lo que se puede considerar que ya disponían de algunos conocimientos previos sobre esta modalidad evaluativa. De forma detallada, se ofrece en las Tablas 1 y 2 la distribución de los estudiantes participantes y de los equipos de trabajo por número de componentes:

Tabla 1. Distribución de los estudiantes participantes

Grupo	Hombres	Mujeres	Total
Grupo A	11	22	33
Grupo B	5	32	37
Grupo C	13	17	30
Total	29	71	100

Tabla 2. Distribución de los equipos de trabajo por número de componentes

Grupo	2 comp.	3 comp.	4 comp.	5 comp.	6 comp.	Total
Grupo A	-	-	3	3	1	7
Grupo B	2	-	-	3	3	8
Grupo C	-	1	4	1	1	7
Total	2	1	7	7	5	22

Sobre los docentes participantes, señalar que fueron 2, un hombre y una mujer, ambos con una experiencia docente en la universidad de unos 10 años. Los docentes compartían la asignatura y cada uno de ellos fue el responsable de 2 de las 4 tareas cuya calificación es objeto de análisis. Concretamente, el Docente A (hombre) guio el proceso de evaluación y calificación de las tareas 1 y 2, por su parte, la Docente B (mujer) hizo lo propio con las tareas 3 y 4.

Instrumento

Para la recogida de información referida a las calificaciones aportadas por parte del docente, de cada estudiante individualmente y la calificación compartida tras el proceso de evaluación conjunta y negociada, se diseñó una

ficha de registro donde se anotaba el nombre del grupo y el número de práctica, en una columna la calificación aportada por el docente, en otras tantas, la calificación individual de los componentes del grupo y en una última columna la nota final acordada tras el proceso de evaluación colaborativa (Figura 3).

Tal y como se ha abordado en el procedimiento, la calificación del docente fue registrada previamente a la reunión presencial de coevaluación, durante la misma se anotaban las valoraciones de cada miembro del grupo y en última instancia se reflejaba la calificación compartida definitiva acordada de la práctica. En todos los casos, el rango utilizado para las calificaciones estaba comprendido entre el 0 como mínimo y el 10 como máximo.

Figura 3. Ficha para la recogida de las calificaciones

CALIFICACIÓN COMPARTIDA									
PRÁCTICA: _____									
GRUPO	DOCENTE	Estudiante 1	Estudiante 2	Estudiante 3	Estudiante 4	Estudiante 5	Estudiante 6	Estudiante 7	Calificación Compartida

Análisis de datos

Una vez registrados los datos, para intentar alcanzar los objetivos de nuestro trabajo, se realizó un análisis estadístico relacionando las tres variables de estudio: La calificación previa del docente a cada práctica, la media de las calificaciones individuales de cada componente de los grupos de trabajo y por último, la calificación compartida final acordada entre estudiantes y docente que sería la nota definitiva de la tarea.

Con la ayuda del software estadístico JASP, se han realizado, en primera instancia, un análisis de los estadísticos descriptivos y posteriormente se ha calculado el coeficiente de correlación de Pearson para comprobar el grado de relación entre las distintas calificaciones

aportadas. De manera complementaria, también se ha realizado la prueba T de Student de muestras relacionadas, con el fin de corroborar si existen diferencias estadísticamente significativas entre las medias pareadas, asimismo, se ha calculado el tamaño del efecto de dichas diferencias usando la d de Cohen. Se decidió utilizar la T de Student con el fin de analizar las diferencias grupo a grupo. Además, se utilizaron las muestras relacionadas debido a que las personas estaban en contacto y podían influenciarse, ya que existía negociación entre ellas. Del mismo modo, los participantes en un grupo también podían formar parte de otro, ya que estudiantes y docentes otorgaban su calificación por separado, pero igualmente decidían conjuntamente la calificación compartida.

Resultados

Análisis descriptivo y correlacional de las calificaciones

Teniendo en consideración que han sido en total 4 prácticas las realizadas por los 100 estudiantes divididos en 22 equipos de trabajo, el número total de tareas calificadas han sido 88. Así, el número de calificaciones analizadas han sido 264: 88 de los 2 docentes (44 de cada),

88 calificaciones compartidas finales y 88 calificaciones que representan la media de 400 valoraciones individuales de los estudiantes participantes. Si analizamos los estadísticos descriptivos de las 3 variables consideradas: La calificación del docente (CD), la calificación media de los componentes del grupo (CG) y la calificación final alcanzada a través de la evaluación colaborativa, a la que denominamos calificación compartida (CC), obtenemos los resultados detallados en la Tabla 3.

Tabla 3. Estadísticos descriptivos por agente evaluador en cada una de las tareas

	Calificación	CD - Docente	CG - Componentes Grupo	CC - Compartida
Tarea 1	M	7.93	8.15	8.18
	DT	0.84	0.59	0.64
	Mín	6	7	7
	Máx	9	7.25	9.25
Tarea 2	M	7.5	7.76	7.8
	DT	0.76	0.66	0.67
	Mín	6	6.5	6.5
	Máx	9	9	9
Tarea 3	M	6.93	7.33	7.17
	DT	1.04	0.91	0.98
	Mín	5	6	5.5
	Máx	9	9	9
Tarea 4	M	7.26	7.68	7.35
	DT	1.75	1.44	1.76
	Mín	3	3.67	3
	Máx	9	9.1	9
Global	M	7.41	7.73	7.63
	DT	1.2	0.98	1.16
	F	2.61	1.91	2.63
	p	.002**	.025*	.002**

Nota: *p<.05, ** p<.01; ***p<.001

Como podría ser previsible atendiendo a otros estudios sobre la participación de los estudiantes en la calificación (Acedo & Ruiz-Cabestre, 2011; Quesada et al., 2017), se puede apreciar que en las 4 tareas las valoraciones más bajas son las otorgadas por los docentes, que de manera global es de 7.41 y está 0.32 puntos por debajo de la calificación de los miembros de los grupos (M=7.73) y 0.22 de las calificaciones compartidas (M=7.63).

Aunque la calificación del docente siempre es menor, se puede comprobar que en las tareas 1

y 2 -recordemos, pertenecientes al Docente A- la calificación compartida final es mayor que la media de las calificaciones individuales de los estudiantes (+0.03 y +0.04), mientras que en las tareas 3 y 4 -Docente B- ocurre exactamente lo contrario (-0.16 y -0.33).

Revisando las puntuaciones máximas y mínimas por cada muestra se observa que son semejantes, siendo en todas ellas la puntuación mínima de los componentes del grupo superior a la de los docentes. Por tareas, destaca la puntuación mínima de la tarea 4, siendo de 3

puntos sobre 10. La desviación típica de las puntuaciones se sitúa entre .59 y 1.76 puntos. Existe una mayor variabilidad en las puntuaciones otorgadas en la tarea 4 (Tabla 3), tanto en la calificación por parte de la docente, como en las otorgadas por los estudiantes y la obtenida en la calificación compartida, obteniendo una desviación típica de entre 1.44 y 1.76 puntos. Esta variabilidad puede ser debida, por una parte, a la mayor complejidad de la última de las tareas, así como por otra parte, al período en el que se realizó la misma, coincidiendo con el final del semestre, que es cuando los estudiantes presentan mayor carga de trabajo. Todo ello pudo favorecer que se llevara a cabo un rendimiento diferente, obteniendo algunos grupos calificaciones elevadas y otros más bajas.

Por otra parte, como se puede comprobar en la Tabla 3 al aplicar el test de Levene (F) se observa que no existe homogeneidad de varianzas entre las puntuaciones otorgadas a cada uno de los equipos de trabajo en las distintas calificaciones analizadas, es decir, las calificaciones de los docentes, la calificación media de los componentes del grupo y la calificación final acordada tras la evaluación colaborativa.

Avanzando un paso más y con el fin de comprobar la fuerza y dirección de la relación existente entre las distintas calificaciones, se ha calculado el coeficiente de correlación de Pearson a partir de las medias de las calificaciones obtenidas en las 4 tareas, cuyos resultados se pueden observar en la Tabla 4.

Tabla 4. Coeficiente de correlación de Pearson

	Correlación r de Pearson	Sig.
Par 1 Media CD - CG	0.893***	< .01
Par 2 Media CD - CC	0.961***	< .01
Par 3 Media CG - CC	0.955***	< .01

Nota: *p<.05, ** p<.01; ***p<.001

Se aprecia con claridad que las distintas calificaciones están fuertemente relacionadas en sentido directo, destacando el gran tamaño del efecto en la variable de la calificación final con la media de los componentes de los grupos ($r=0.955$) y en la calificación del docente con la calificación final ($r=0.961$). Por lo tanto, se puede afirmar, apreciando las evidencias, que existe una estrecha correlación positiva entre las distintas calificaciones, aumentando y disminuyendo simultáneamente en función de las tareas evaluadas y calificadas. Utilizando otras palabras, se demuestra que cada agente evaluador propone de manera proporcionada sus calificaciones dependiendo del propio objeto de evaluación.

Si segmentamos las calificaciones por tareas y por docentes, obtenemos los siguientes resultados al calcular el coeficiente de correlación de Pearson (Tabla 5 y Tabla 6).

Tabla 5. Coeficiente de correlación de Pearson desagregado por docente

	Docente A r de Pearson	Docente B r de Pearson
Par 1 Media CD - CG	0.831***	0.909***
Par 2 Media CD - CC	0.892***	0.981***
Par 3 Media CG - CC	0.948***	0.956***

Nota: *p<.05, ** p<.01; ***p<.00

Tabla 6. Coeficiente de correlación de Pearson desagregado por tarea

	Tarea 1 r de Pearson	Tarea 2 r de Pearson	Tarea 3 r de Pearson	Tarea 4 r de Pearson
Par 1 CD - CG	0.837***	0.805***	0.850***	0.936***
Par 2 CD - CC	0.924***	0.845***	0.951***	0.993***
Par 3 CG - CC	0.919***	0.964***	0.933***	0.965***

Nota: *p<.05, ** p<.01; ***p<.00

En todas las tareas existe una alta correlación entre las tres puntuaciones obtenidas. En la tarea 4 se aprecia una muy fuerte correlación entre las distintas puntuaciones ofrecidas, seguida de la práctica 3. En coherencia con ello,

revisando las puntuaciones de los docentes, vemos que se obtiene una correlación más fuerte en el caso de la Docente B, que se encargaba de la corrección de las tareas 3 y 4.

Análisis de las diferencias entre las calificaciones

Para completar el estudio, se aplicó la prueba T de Student de muestras relacionadas para comparar las medias de las calificaciones de las 4 tareas de evaluación. Las puntuaciones medias obtenidas son relativamente elevadas (Tabla 3). Tal y como se puede apreciar en la Tabla 7, se han hallado diferencias estadísticamente significativas con un tamaño del efecto mediano entre las calificaciones otorgadas por el docente tanto con las

calificaciones de los estudiantes componentes de los grupos ($M(CD)=7.41$; $M(CG)=7.73$; $p<.05$; $d=0.545$), como con la calificación compartida ($M(CD)=7.41$; $M(CC)=7.63$; $p<.05$), con un tamaño del efecto de mediano a grande ($d=0.670$). No obstante, no se presentan diferencias estadísticamente significativas entre la variable de la calificación compartida y las calificaciones aportadas por el equipo de estudiantes ($M(CC)=7.63$; $M(CG)=7.73$; $p>.05$), en este caso el tamaño del efecto es muy pequeño ($d=0.148$).

Tabla 7. T de Student de muestras relacionadas en la media obtenida en las cuatro tareas

	t	Sig. (bilateral)	d de Cohen	IC 95% d de Cohen	
				Inferior	Superior
Media CD - CG	-5.14	< .01***	0.545	-0.775	-0.324
Media CD - CC	-6.25	< .01***	0.670	-0.896	-0.433
Media CG - CC	1.55	.12	0.148	-0.045	0.378

Nota: * $p<.05$, ** $p<.01$; *** $p<.001$

Revisando los datos desagregados, al considerar las diferencias en cada una de las tareas (Tabla 8), podemos observar que se hallan diferencias estadísticamente significativas en 3 de las 4 tareas entre la calificación del docente y las calificaciones de los componentes de los grupos ($p<.05$). En la tarea 4, si bien la media de la calificación del docente ($M=7.26$) sigue siendo inferior a la

media de la calificación del grupo de estudiantes ($M=7.68$), de manera opuesta a los datos globales, no existen diferencias estadísticamente significativas entre ambas calificaciones, no obstante, sí aparece un tamaño del efecto de la diferencia mediano ($p=.412$), al igual que ocurre en las tareas 1, 2 ($d=0.458$ y 0.586 respectivamente) y de mediano a grande en la tarea 3 ($d=0.740$).

Tabla 8. T de Student de muestras relacionadas en cada una de las tareas

	t	Gl	Sig. (bilateral)	d de Cohen	IC 95% d de Cohen	
					Inferior	Superior
TAREA 1						
Par 1 CD - CG	2.15	21	.04*	0.458	-0.893	-0.013
Par 2 CD - CC	-3.40	21	.003**	0.725	-1.189	-0.247
Par 3 CG - CC	-0.64	21	.53	0.137	-0.555	0.285
TAREA 2						
Par 1 CD - CG	-2.75	21	.012**	0.586	-1.033	-0.126
Par 2 CD - CC	-3.42	21	.003**	0.728	-1.193	-0.250
Par 3 CG - CC	-0.83	21	.417	0.176	-0.596	0.247
TAREA 3						
Par 1 CD - CG	-3.39	21	.003**	0.722	-1.187	-0.245
Par 2 CD - CC	-3.47	21	.002**	0.740	-1.206	-0.260
Par 3 CG - CC	2.13	21	.046*	0.453	0.009	0.888
TAREA 4						
Par 1 CD - CG	-1.89	20	.074	0.412	-0.853	0.039
Par 2 CD - CC	-2.22	21	.038*	0.473	-0.910	-0.027
Par 3 CG - CC	1.37	20	.187	0.298	-0.143	0.732

Nota: * $p<.05$, ** $p<.01$

En todas las tareas existen diferencias estadísticamente significativas entre la puntuación que hubiera otorgado el docente en solitario (con puntuaciones medias comprendidas entre 6.93 y 7.93) y la puntuación obtenida en la calificación consensuada ($p < .05$) (con puntuaciones medias entre 7.17 y 8.18 puntos). Estas diferencias tienen un tamaño del efecto de mediano a grande (d alrededor de 0.7) en las tres primeras tareas, y un tamaño mediano en la última tarea ($d = -0.473$).

A diferencia de lo que se aprecia en los datos agregados, existe una tarea, la 3, en la que también existen diferencias estadísticamente significativas ($p < .05$), con un tamaño del efecto mediano ($d = 0.453$) entre la puntuación de los miembros del equipo ($M = 7.33$) y la calificación compartida de la tarea ($M = 7.17$).

En último lugar, si realizamos el contraste estadístico teniendo en cuenta el docente que se encargaba de la evaluación de cada tarea, se pueden apreciar en la Tabla 9 diferencias entre ambos. De este modo, al calcular la *T* de Student para muestras relacionadas en función del docente, puede apreciarse que existen diferencias estadísticamente significativas entre la media de cada uno de los docentes y la calificación de los estudiantes con un tamaño del efecto mediano en cada caso (Docente A:

$M(CD) = 7.72$; $M(CG) = 7.96$; $p = .001$; $d = 0.526$; Docente B: $M(CD) = 7.09$; $M(CG) = 7.5$; $p < .001$; $d = 0.572$), asimismo, en ambos profesores existen diferencias estadísticamente significativas entre la calificación del docente y la calificación derivada de la evaluación colaborativa, con un tamaño del efecto mediano en el caso del Docente B y de mediano a grande en el caso del Docente A (Docente A: $M(CD) = 7.72$; $M(CC) = 7.99$; $p < .001$; $d = 0.731$; Docente B: $M(CD) = 7.09$; $M(CG) = 7.26$, $p < .001$; $d = 0.607$). En ambos casos la media de la evaluación colaborativa es superior a la calificación que hubiera otorgado el docente en solitario.

Asimismo, en relación entre las puntuaciones otorgadas por los estudiantes y las derivadas de la calificación compartida, en el Docente A, las puntuaciones medias ofrecidas por el equipo de estudiantes ($M = 7.96$) son semejantes a las de la calificación final ($M = 7.99$;) y no aparecen diferencias estadísticamente significativas entre ambas ($p = .317$; $d = 0.153$), mientras que, en el caso del Docente B, sí que aparecen diferencias estadísticamente significativas entre ambas medias, siendo inferior la puntuación obtenida en la calificación compartida, con un tamaño del efecto de pequeño a mediano ($M(CG) = 7.5$; $M(CC) = 7.26$; $p = .02$; $d = 0.376$).

Tabla 9. *T* de Student de muestras relacionadas en función del docente

					IC 95% d de Cohen	
	t	Gl	Sig. (bilateral)	d de Cohen	Inferior	Superior
DOCENTE 1						
Par 1 Media CD - CG	-3.49	43	.001**	-0.526	-0.839	-0.208
Par 2 Media CD – CC	-4.85	43	<.001***	-0.731	-1.061	-0.395
Par 3 Media CG - CC	-1.01	43	.317	-0.153	-0.449	0.146
DOCENTE 2						
Par 1 Media CD - CG	-3.75	42	<.001***	-0.572	-0.892	-0.247
Par 2 Media CD – CC	-4.02	43	<.001***	-0.607	-0.926	-0.282
Par 3 Media CG - CC	2.47	42	.02*	0.376	0.064	0.683

Nota: * $p < .05$, ** $p < .01$, *** $p < .001$

Discusión y conclusiones

Con este análisis se ha corroborado que las calificaciones compartidas en la evaluación colaborativa y las calificaciones del docente están fuertemente correlacionadas en sentido directo. No obstante, aunque correlacionadas,

se ha detectado la existencia de diferencias estadísticamente significativas entre ellas con un tamaño del efecto entre mediano y grande. Por tanto, las calificaciones son más altas si se realiza una calificación compartida y

consensuada que si únicamente calificara el docente.

De hecho, se ha evidenciado que las calificaciones derivadas de la evaluación colaborativa han sido superiores en +0,32 puntos, en términos de medias generales, respecto a la calificación que inicialmente habría aportado el docente el solitario. Estas calificaciones compartidas han sido -0,10 puntos menores respecto a las que aportarían, de media, los estudiantes respecto a sus propios desempeños en las tareas, algo que se puede relacionar con otros estudios sobre las calificaciones en la autoevaluación que suelen ser las más elevadas de las modalidades participativas (Acedo & Ruiz-Cabestre, 2011; Quesada et al., 2016). No obstante, y de forma contraria al caso anterior, no se han hallado en nuestros resultados la existencia de diferencias estadísticamente significativas entre las calificaciones compartidas y las calificaciones medias de los grupos de estudiantes con un tamaño del efecto muy pequeño.

Al desagregar los datos, se ha constatado que solo en una tarea, la 3, hay diferencias estadísticamente significativas entre la calificación que se hubieran asignado los estudiantes de los grupos y la que finalmente obtuvieron en la calificación compartida con un tamaño del efecto mediano. Esta circunstancia es la causante de que, igualmente, aparezcan diferencias estadísticamente significativas entre la media de los componentes y las calificaciones finales en la Docente B con un tamaño del efecto de pequeño a mediano. En este punto, es importante reflexionar sobre este hecho, y es que, esta fue la primera práctica que evaluó esta docente con estos grupos de estudiantes. Como se comentó anteriormente, el alumnado ya había cursado otra asignatura con el Docente A y tenía experiencia con él en la evaluación colaborativa y la calificación compartida. Al desaparecer nuevamente las diferencias en la siguiente tarea, la 4, siendo además la única donde tampoco existen diferencias estadísticamente significativas entre la calificación media del equipo y la calificación de la docente, podría suponerse la existencia de un proceso de aprendizaje durante

el desarrollo de la evaluación colaborativa, y cómo, en función del docente, quizá los alumnos se van adaptando a su forma de valorar los trabajos. Las lecturas en este sentido pueden abrir nuevas vías de investigación basadas en el estudio de los aprendizajes que subyacen al proceso, las interacciones, relaciones emocionales y personales, así como las competencias del profesorado para la gestión de grupos cuando se desarrolla este tipo de evaluaciones.

A pesar de las limitaciones propias respecto a la variedad de los participantes y a la presencia de las diferencias estadísticamente significativas, se considera que los resultados pueden ser igualmente interesantes para reflexionar sobre sus causas. Parece plausible que las consecuencias académicas, sociales y económicas de las calificaciones no pueden obviarse y que, a su vez, deberían tenerse en consideración para atisbar la influencia real que puede tener la participación de los estudiantes en procesos de evaluación, al menos respecto a la educación formal en las instituciones universitarias actuales. ¿Hasta qué punto se puede ajustar unas calificaciones cuando lo que se gana o lo que se pierde trasciende en mucho a las consecuencias cognitivas del aprendizaje? Tendríamos que preguntarnos si, a pesar de los esfuerzos por sistematizar la evaluación y las propias prácticas de alfabetización evaluativa, realmente se podrían ignorar todas las connotaciones que las calificaciones arrastran consigo. Aunque sea fuente de problema, tampoco consideramos que obviarla o eliminarla de la participación de los estudiantes sea lo más coherente o deseable, sobre todo desde los planteamientos de extrapolar la lógica del aprendizaje activo como vía de construcción del conocimiento efectivo y significativo también al ámbito de la evaluación. En este sentido, estamos de acuerdo con McArthur (2019) cuando plantea que una posible solución sería liberar a la evaluación de esquemas rígidos de calificación numérica ya que esos niveles de precisión y diferenciación ni se corresponden con las realidades complejas de la mayoría de tareas en Educación Superior y a su vez, “desvía la

atención de lo que realmente importa, que es la aplicación social de tal conocimiento para fomentar un mayor bienestar individual y social” (p.132).

Por otro lado, es importante volver a subrayar que, en esta experiencia, a diferencia de estudios anteriores (Gómez & Quesada, 2017; Quesada et al., 2019), los docentes han calificado previamente las prácticas y los estudiantes han valorado individualmente su desempeño, por tanto, al haber un trabajo previo de valoración en solitario, la calificación no ha surgido únicamente de la reflexión conjunta y la interacción dialógica. Habría, pues, que tener en consideración cómo estas modificaciones imprescindibles en el diseño, eminentemente cuantitativo, para una recogida de información que permitiera obtener los datos necesarios que nos facilitaran cumplir los objetivos de la investigación, han podido influir -positiva o negativamente- o han acarreado una serie de limitaciones y sesgos necesarios a valorar seriamente en futuras experiencias similares.

Insistiendo en lo anterior, es necesario recalcar que en esta experiencia el docente había tenido que reflexionar y anotar previamente al proceso dialógico la calificación de cada una de las tareas de evaluación. Por lo tanto, podría parecer factible que, en uno u otro sentido, esto hubiera podido condicionar el proceso o incluso que directamente se tendiera a esa calificación, aunque esto último, al menos en nuestro caso, se ha demostrado que no fue así. Igualmente, los docentes percibieron que, por el motivo anterior y también en función del diálogo previo a la calificación -sobre la evaluación y apreciación colaborativa de la tarea-, en ocasiones parecía que los estudiantes pretendían adivinar la calificación del docente más que realizar un verdadero acto de auto-calificación consecuente.

Complementariamente, y al aportar cada miembro del grupo una calificación individual, el diálogo sobre la calificación compartida en ocasiones no era del todo fluido entre los participantes, proponiendo en algunos casos realizar directamente la media para zanjar el

asunto. Esto sí que se ha visto reflejado en los resultados -específicamente con el Docente A se dio de manera evidente-, pues el hecho de proponer la media entre las calificaciones individuales de los estudiantes en la calificación compartida podría explicar la ausencia de diferencias estadísticamente significativas entre las calificaciones finales y las calificaciones de los miembros de los grupos de trabajo con el mencionado docente y, a partir de ahí, de manera general.

En conclusión, consideramos que además de la alta correlación de las calificaciones obtenidas en este estudio concreto, beneficios formativos como el aumento del esfuerzo, motivación e implicación (Dochy et al., 1999), el fomento de la reflexión de la propia actuación (Gómez & Quesada, 2017; Knight & Yorke, 2003; Quesada et al. 2019), y la mejora de la autoconfianza del alumnado participante (Boud & Falchikov, 2006) son razones suficientes que avalan y hacen recomendable no solo utilizar esta modalidad evaluativa, sino también continuar con su indagación científica. Así, otro aspecto clave a profundizar en base a los resultados es el papel que desempeña en el proceso evaluativo la existencia de una calificación compartida y la actitud de los implicados en el proceso. Esto se podrá comprobar en futuros estudios mediante la grabación de las sesiones de evaluación compartida y negociada, ya sean estas presenciales o virtuales mediante herramientas de videoconferencia.

Por último, señalamos como necesario que en los diseños de las evaluaciones colaborativas, además de los elementos técnicos y formales, se analicen y consideren las relaciones de poder existentes y se incluyan estrategias para reducir estos desequilibrios, máxime cuando se finalice mediante una calificación compartida. La importancia de reconocer y trabajar con las complejidades de la evaluación no debe ser subestimado (Cooper, 2017), especialmente cuando se incluye un elemento tan controvertido y con tantas repercusiones como la calificación en los contextos de la educación universitaria.

Referencias

- Acedo, M.A., & Ruiz-Cabestre, F.J. (2011). Una experiencia sobre la evaluación autónoma o participativa: autoevaluación y evaluación por los compañeros. *Arbor*, 187, 183-188.
<https://doi.org/10.3989/arbor.2011.Extra-3n3142>
- Álvarez, J.M. (2001). *Evaluar para conocer, examinar para excluir*. Morata.
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4ª Edición). McGraw-Hill-SRHE & Open University Press.
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399-413.
<https://doi.org/10.1080/02602930600679050>
- Boud, D., & Soler, R. (2015). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400-4013.
<https://doi.org/10.1080/02602938.2015.1018133>
- Bretones, A. (2008). Participación del alumnado de Educación Superior en su evaluación. *Revista de Educación*, 347, 181-202.
<http://www.educacionyfp.gob.es/dam/jcr:07d3f344-a3f9-4124-b4d1-c2b2d9f8bb27/re34709-pdf.pdf>
- Canabal, C., & Margalef, L. (2017). La retroalimentación: La clave para una evaluación orientada al aprendizaje. *Profesorado. Revista de currículum y formación del profesorado*, 21(2), 149-170.
<http://hdl.handle.net/10481/47669>
- Carless, (2007). Learning-oriented assessment: conceptual bases and practical implications. *Innovations in Education and Teaching international*, 57(66), 57-66.
<https://doi.org/10.1080/14703290601081332>
- Casanova, M.A. (2011). Evaluación para la Inclusión Educativa. *Revista Iberoamericana de Evaluación Educativa*, 4(1), 79-89.
<http://www.rinace.net/riee/numeros/vol4-num1/art4.pdf>
- Cooper, S. (2017). A collaborative assessment of students' placement learning. *Assessment & Evaluation in Higher Education*, 42(1), 61-76.
<https://doi.org/10.1080/02602938.2015.1083093>
- Crisp, G.T. (2012) Integrative assessment: reframing assessment practice for current and future learning. *Assessment & Evaluation in Higher Education*, 37(1), 33-43.
<https://doi.org/10.1080/02602938.2010.494234>
- Deeley, S. (2014). Summative co-assessment: A deep learning approach to enhancing employability skills and attributes. *Active Learning in Higher Education*, 15(1) 39-51.
<https://doi.org/10.1177/1469787413514649>
- Dochy, F., Segers M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331-350.
<https://doi.org/10.1080/03075079912331379935>
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287-322. <https://doi.org/10.2307/1170785>
- García, E., Gallego, B. & Gómez, M.A. (2015). Feedback and self-regulated learning: How feedback can contribute to increase students' autonomy as learners. En M. Peris-Ortiz & J.M. Merigó-Lindahl. *Sustainable Learning in Higher Education. Developing Competencies for the Global Marketplace* (113-130). Springer. https://doi.org/10.1007/978-3-319-10804-9_9
- Gallego, B., Quesada, V., Gómez, M.A., & Cubero, J. (2017). La evaluación y retroalimentación electrónica entre iguales para la autorregulación y el aprendizaje estratégico en la universidad: la percepción del alumnado. *REDU. Revista de Docencia Universitaria* 15 (1): 127-146.
<https://doi.org/10.4995/REDU.2017.5991>
- Gómez, M.A., & Quesada, V. (2017). Coevaluación o Evaluación Compartida en el Contexto Universitario: La Percepción del Alumnado de Primer Curso. *Revista Iberoamericana de Evaluación Educativa*, 10 (2), 9-30.
<https://doi.org/10.15366/riee2017.10.2.001>

- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
<https://doi.org/10.3102/003465430298487>
- Hayward, L. (2015). Assessment is learning: the preposition vanishes. *Assessment in Education: Principles, Policy & Practice*, 22(1), 27-43.
<https://doi.org/10.1080/0969594X.2014.984656>
- Hernández, F., & Maquilón, J.J. (2010). Introducción a los diseños de investigación educativa. En S. Nieto (Ed.), *Principios, métodos y técnicas esenciales para la investigación educativa* (109-126). Dykinson.
- Hidalgo, N., & Murillo, F.J. (2016). Evaluación de Estudiantes para la Justicia Social. Propuesta de un Modelo. *Revista Internacional de Educación para la Justicia Social (RIEJS)*, 5(2), 159-179.
<https://doi.org/10.15366/riejs2016.5.2.008>
- Ibarra, M.S., & Rodríguez, G. (2019). Evaluación como aprendizaje. En J. Paricio, A. Fernández & I. Fernández (Eds.), *Cartografía de la buena docencia universitaria. Un marco para el desarrollo del profesorado basado en la investigación* (175-196). Narcea.
- Knight, P. T., & Yorke, M. (2003). *Assessment, learning and employability*. SRHE/Open University Press/McGraw-Hill Education.
- Kurt, M. (2014). Collaborative Assessment: Fostering ownership in assessment. *Education*, 134(3), 332-339.
- Leach, L., Neutze, G., & Zepke, N. (2001). Assessment and Empowerment: Some critical questions. *Assessment and Evaluation in Higher Education*, 26(4), 293-305.
<https://doi.org/10.1080/02602930120063457>
- López, V.M. (2009). *Evaluación formativa y compartida en Educación Superior. Propuestas, técnicas, instrumentos y experiencias*. Narcea.
- López, V.M. (2012). Evaluación formativa y compartida en la universidad: clarificación de conceptos y propuestas de intervención desde la Red Interuniversitaria de Evaluación Formativa. *Psychology, Society & Education*, 4(1), 117-130.
<https://doi.org/10.25115/psyse.v4i1.485>
- López, V., & Sicilia, A. (2017). Formative and shared assessment in higher education. Lessons learned and challenges for the future. *Assessment & Evaluation in Higher Education*, 42(1), 77-97.
<https://doi.org/10.1080/02602938.2015.1083535>
- McArthur, J. (2019). *La evaluación: una cuestión de justicia social. Perspectiva crítica y prácticas adecuadas*. Narcea.
- Murillo, F.J., & Hidalgo, N. (2015). Enfoques Fundamentales de la Evaluación de Estudiantes para la Justicia Social. *Revista Iberoamericana de Evaluación Educativa*, 8(1), 43-61.
<https://revistas.uam.es/index.php/riece/article/view/2975>
- Nicol, D. (2010). From monologue to dialogue: improving written feedback in mass higher education. *Assessment and Evaluation in Higher Education*, 35(5), 501-517.
<https://doi.org/10.1080/02602931003786559>
- Nicol, D. (2013). Peer review: putting feedback processes in students' hands. *Perspectives on Pedagogy and Practice*, 4, 111-123.
http://uir.ulster.ac.uk/26926/1/PERSPECTIVE_ON_PEDAGOGY_AND_PRACTICE.pdf
- Penuel, W.R., & Shepard, L.A. (2016). Social models of learning and assessment. En A.A. Rupp & J.P. Leighton (Eds.), *The handbook of cognition and assessment: Frameworks, methodologies, and applications* (146-173). John Wiley & Sons.
- Quesada, V., García, E., & Gómez, M.A. (2017). Student participation in assessment processes: A way forward. En E. Cano & G. Ion (Eds.), *Innovative practices for higher education assessment and measurement* (228-249). Hershey: IGI Global.
<https://doi.org/10.4018/978-1-5225-0531-0.ch012>
- Quesada, V., Gómez, M.A., Gallego, B., & Cubero, J. (2019). Should I use co-assessment in higher education? Pros and cons from teachers and students' perspectives. *Assessment & Evaluation in Higher Education*.
<https://doi.org/10.1080/02602938.2018.1531970>

- Quesada, V., Rodríguez, G., & Ibarra, M.S. (2016). What are we missing? Spanish lecturers' perceptions of their assessment practices. *Innovations in Education and Teaching International*, 53(1), 48-59. <https://doi.org/10.1080/14703297.2014.930353>
- Quesada, V., Rodríguez, G., & Ibarra, M.S. (2017). Planificación e innovación de la evaluación en educación superior: la perspectiva del profesorado. *Revista de investigación educativa*, 35(1), 53-70. <https://doi.org/10.6018/rie.35.1.239261>
- Rodríguez, G., & Ibarra, M.S. (2015). Assessment as Learning and Empowerment: Beyond Sustainable Learning in Higher Education. En M. Peris-Ortiz & J.M. Merigó-Lindahl (Eds.), *Sustainable learning in higher education, innovation, technology, and knowledge management* (1-20). Springer-Verlag. https://doi.org/10.1007/978-3-319-10804-9_1
- Santiuste, V., & Arranz, M.L. (2009). Nuevas perspectivas en el concepto de evaluación. *Revista de Educación*, 350, 463-476. http://www.revistaeducacion.educacion.es/re350/re350_20.pdf
- Smith C.D., Worsfold, K., Davies, L., Fisher, R., & McPhail, R. (2013). Assessment literacy and student learning: the case for explicitly developing students 'assessment literacy'. *Assessment & Evaluation in Higher Education*, 38(1), 44-60. <https://doi.org/10.1080/02602938.2011.598636>
- Swaffield, S. (2011). Getting to the heart of authentic Assessment for Learning. *Assessment in Education: Principles, Policy & Practice*, 18(4), 433-449. <https://doi.org/10.1080/0969594X.2011.582838>
- Topping, K.J. (2003). Self and peer assessment in school and university: reliability, validity and utility. En M. Segers, F. Dochy & E. Cascallar (Eds), *Optimizing new modes of assessment: in search of qualities and standards* (55-87). Springer. https://doi.org/10.1007/0-306-48125-1_4
- Van der Bergh, V., Mortelmans, D., Spooren, P., Van Petegem, P., Gijbels, D., & Vanthournout, G. (2006). New assessment modes within project-based education - The stakeholders. *Studies in Educational Evaluation*, 32, 345-368. <https://doi.org/10.1016/j.stueduc.2006.10.005>
- Wood, P., y Smith, J. (2018). *Investigar en educación. Conceptos básicos y metodología para desarrollar proyectos de investigación*. Narcea.

Authors / Autores

Gómez-Ruiz, M.A. (miguel.gomez@uca.es)  0000-0003-2418-9525

Profesor Ayudante Doctor en el Área de Didáctica y Organización Escolar del Departamento de Didáctica de la Universidad de Cádiz. Miembro del Grupo de Investigación EVALfor (Evaluación en Contextos Formativos, SEJ-509). Investigador de la Cátedra UNESCO en Evaluación, Innovación y Excelencia de la Universidad de Cádiz. Sus principales líneas de investigación sobre evaluación se centran en la evaluación orientada al aprendizaje, la participación en procesos de evaluación y la evaluación dialógica y compartida.

Quesada-Serra, V. (victoria.quesada@uib.eu)  0000-0002-8881-2358

Profesora Contratada Doctor en el Departamento de Pedagogía y Didácticas Específicas de la Universitat de les Illes Balears. Miembro del Grupo de Investigación GIFES (Investigación y Formación Educativa y Social). Sus principales líneas de investigación sobre evaluación se centran en la evaluación orientada al aprendizaje, la participación en procesos de evaluación y la evaluación dialógica y compartida.



Revista ELecciónica de **I**nvestigación y **E**Valuación **E**ducativa
E-Journal of Educational Research, Assessment and Evaluation
[ISSN: 1134-4032]



Esta obra tiene [licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional](https://creativecommons.org/licenses/by-nc/4.0/).
This work is under a [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/).