

Una experiencia de evaluación por pares y autoevaluación combinada con finalidad formativa y sumativa.

An experience of hybrid peer and self-assessment with a formative and summative purpose

Juan Andrés González^a y Jesús Mengual^b

^aUniversitat Politècnica de València, Escuela Politécnica Superior de Gandia, juagonr1@hma.upv.es, 

^bUniversitat Politècnica de València, Escuela Politécnica Superior de Gandia, jemencu@hma.upv.es, 

How to cite: González, J.A. y Mengual, J. 2023. Una experiencia de evaluación por pares y autoevaluación combinada con finalidad formativa y sumativa. En libro de actas: *IX Congreso de Innovación Educativa y Docencia en Red*. Valencia, 13 - 14 de julio de 2023. Doi: <https://doi.org/10.4995/INRED2023.2023.16672>

Abstract

This study analyses the incorporation of students in the assessment process. Firstly, with a summative purpose, the reliability and correlation between lecturers' and students' qualifications have been evaluated. In that sense, the experience of student participation in the assessment of oral presentations during three academic years has been analysed. To minimize the lack of confidence in the impartiality of student marks, the use of a standardization algorithm has been established in order to reduce discrepancies between marks of students and teachers. Secondly, with a formative purpose, the impact of a combined use of self-evaluation and peer evaluation has been analysed. This combination implies a self-reflexive process that contributes to the awareness about the quality of their own work, as well as to contain an overestimation of students towards their peers. In addition, an immediate improvement in academic results of those who, previously, had qualified their peers, has been observed. Finally, through an opinion survey, students seem to be favourable to their participation in the assessment process but not without certain resistance.

Keywords: *peer evaluation, self-assessment, student perception, standardization, self-reflexive learning, critical thinking.*

Resumen

Este trabajo analiza la incorporación del alumnado en el proceso de evaluación. En primer lugar, con finalidad sumativa, se ha evaluado la fiabilidad y correlación de las calificaciones del profesorado y del alumnado. Para ello, se analiza la experiencia participativa del alumnado en la evaluación de exposiciones orales de trabajos académicos durante tres cursos académicos. Para minimizar la falta de confianza sobre la imparcialidad de las calificaciones del alumnado, se ha establecido y propuesto el uso de un algoritmo de estandarización de calificaciones que permita reducir las discrepancias entre los grupos de alumnos y profesores. En segundo lugar, con carácter formativo, se ha analizado el impacto del uso combinado de la autoevaluación y evaluación por pares, observándose que éste implica un proceso autorreflexivo que contribuye a la toma de conciencia sobre la calidad

del trabajo propio, así como a una contención de la sobrevaloración del estudiante hacia sus pares. Además, se ha observado una mejora inmediata en el resultado académico de aquellos estudiantes que, previamente, han participado como evaluadores de sus pares. Por último, mediante una encuesta de opinión, se detecta una valoración general del alumnado favorable a su participación en el proceso de evaluación pero no exenta de resistencias.

Palabras clave: *evaluación por pares, autoevaluación, percepción del alumnado, estandarización, aprendizaje autorreflexivo, pensamiento crítico.*

1. Introducción.

Entre las principales funciones de la Educación Superior se encuentra la necesidad de ayudar a los estudiantes a desarrollar competencias de carácter transversal como, entre otras, la comprensión e integración, el pensamiento crítico, la comunicación, la resolución de problemas y el aprendizaje autónomo. Este proceso debe desarrollarse de forma continua y constituir un aprendizaje permanente que se extienda a su futura vida profesional. Sin embargo, los métodos tradicionales de evaluación, generalmente, no suelen representar una vía que facilite el pensamiento crítico y reflexivo, así como la autoevaluación del alumnado (Dochy y Moerkerke, 1997).

En este sentido, algunos métodos de evaluación alternativos, como la autoevaluación, evaluación por pares y la evaluación colaborativa, potencian la adquisición de estas competencias, además de constituir procesos en los que el alumnado es protagonista de su propio aprendizaje. Así, la autoevaluación hace referencia a involucrar al alumno en la elaboración de juicios sobre su propio aprendizaje. Por otra parte, la evaluación por pares define el proceso donde los alumnos evalúan a sus iguales a través de un proceso, generalmente, guiado y previamente diseñado. Por último, la evaluación colaborativa consiste en la participación conjunta del alumnado y profesorado en el proceso de evaluación, que posibilita al alumno participar en su propia evaluación, manteniéndose un control del profesorado sobre la evaluación final (Falchikov, 1986).

En definitiva, el uso de estos métodos logra en el alumnado acentuar su rendimiento, mejorar la comprensión de contenidos, desarrollar habilidades interpersonales, afianzar la confianza y la capacidad de autoevaluación y para evaluar a otros (Ibarra et al., 2012).

Sin embargo, a pesar de sus evidentes beneficios, su uso todavía está poco extendido en algunos ámbitos. Fundamentalmente, ello es atribuido a la existencia de una percepción de determinados riesgos que actúa como barrera para la implementación de estos métodos de evaluación. Estas barreras son percibidas como limitaciones, tanto por parte del profesorado como por parte del alumnado. En la Tabla 1 se recogen algunas de estas limitaciones.

En este sentido, el rechazo principal por parte del profesorado está asociado a las dudas que genera la confianza en la capacidad del alumnado para actuar de un modo justo, ajeno a otras motivaciones de carácter personal. Curiosamente, del mismo modo, una parte del alumnado también tiende a ser reactivo a estos métodos cuando implican la participación directa de otros compañeros, bien por su temor a la falta de imparcialidad, o bien por no sentirse cómodos emitiendo juicios de valor sobre sus propios compañeros, especialmente, si ello se percibe como un riesgo de alteración de sus relaciones personales.

Tabla 1. Riesgos y barreras para la aplicación de la autoevaluación y evaluación por pares.

Implicación	Percibidos por el profesorado	Percibidos por el alumnado
Institucional	Reacciones de colegas y evaluadores externos.	Creencia que la evaluación corresponde al profesor.
Dedicación	Demanda excesiva de tiempo y esfuerzo durante el diseño y aplicación.	Pérdida de tiempo.
Objetividad	Falta de robustez en la aplicación del método.	Sensación que la evaluación sumativa es más objetiva.
Confianza	Dudas sobre la imparcialidad de los estudiantes.	Temor ante la imparcialidad del resto de compañeros.
Otros	Temor a la falta de fiabilidad.	Afección negativa a la relación personal cuando el resultado no es favorable.

Fuente: Elaboración propia (adaptado de Romá et al., 2022)

2. Objetivos.

El objetivo general de este trabajo es explorar los riesgos y beneficios de la aplicación y uso combinado de la autoevaluación y evaluación por pares, con carácter formativo y sumativo, durante la presentación de trabajos académicos en el ámbito universitario. Para lograr dicho objetivo, como objetivos específicos, se propone analizar la correspondencia entre las calificaciones del alumnado y profesorado, implementar un algoritmo de cálculo que reduzca la diferencia entre las calificaciones de pares y profesorado, analizar el impacto sobre la mejora inmediata en el rendimiento y detectar las percepciones que el alumnado tiene ante este tipo de evaluación.

3. Desarrollo de la innovación.

A pesar de los importantes beneficios que la participación del alumnado en los procesos de evaluación conlleva para su aprendizaje, todavía siguen existiendo un conjunto de percepciones negativas que limitan su uso y generan cierta resistencia por parte del profesorado y alumnado. Probablemente, la principal preocupación que despierta en el profesorado es la fiabilidad de las calificaciones emitidas por los alumnos hacia sus compañeros. En concreto, preocupa la falta de similitud entre estas calificaciones y las emitidas por el profesor. Por su parte, algunos estudiantes también se muestran reacios a ser calificados por sus compañeros.

Por ello, mejorar la confianza en la capacidad del alumnado para actuar como evaluadores es clave para reducir las reticencias existentes hacia esta metodología de evaluación. En este sentido, es necesario medir la correlación entre las calificaciones del alumnado y profesor, así como buscar estrategias que contribuyan a evitar estos problemas. De este modo, si se mejorase esta discrepancia, se reduciría la reticencia hacia una metodología que aporta beneficios de carácter formativo.

En este sentido, el presente estudio se ha desarrollado en el marco de la asignatura *Gestión de Materiales y Energía* (4º curso del Grado en Ciencias Ambientales de la Universitat Politècnica de València). La evaluación con participación del alumnado se ha planteado con carácter sumativo. En este sentido, si bien es ampliamente aceptado el potencial de la evaluación por pares, su inclusión en las calificaciones es más polémico (Magin y Helmore, 2001). No obstante, si bien existen argumentos a favor y en contra de su incorporación con carácter sumativo (White, 2009), consideramos que su exclusión reduciría el grado de responsabilidad adquirido por los estudiantes durante su rol de evaluador.

Durante la fase de diseño, se ha considerado adecuado la selección de una actividad referida a contenidos de carácter menos técnico y/o complejo. Como actividad objeto de la evaluación se ha seleccionado el trabajo académico desarrollado en el bloque de *Gestión de Materiales*, que constituye un 20% de la nota. La actividad seleccionada permite, además, un desarrollo de dificultad similar para todo el alumnado.

El trabajo académico se ha realizado de forma autónoma mediante un trabajo grupal con grupos de pequeño tamaño (3 alumnos). El profesorado ha ejercido una función de seguimiento durante su desarrollo. Con suficiente antelación, el profesorado ha facilitado una guía con los criterios que debían ser considerados durante el desarrollo del trabajo. Posteriormente, todos los trabajos han sido expuestos públicamente en el aula a través de presentaciones de 15 minutos de duración. Atarés et al. (2021) señalan que la valoración general como síntesis de aspectos fundamentales y formales, a través de una calificación global, la naturaleza de las fuentes de información y la utilización de material gráfico constituyen los aspectos en los que la coevaluación resultaría más fiable. En este sentido, durante la exposición de los trabajos, los alumnos asistentes han tomado notas con la finalidad de evaluar los trabajos de sus compañeros en dos ámbitos principales: (i) contenidos (selección de contenidos, dominio del tema, orden y coherencia) y (ii) presentación (uso de materiales de carácter audio-visual, lenguaje y comunicación, puesta en escena). Tras cada una de las exposiciones, los alumnos han compartido sus valoraciones con otros compañeros, emitiendo, finalmente, una calificación del trabajo con carácter individual y anónimo, señalándose el grupo de pertenencia del evaluador. Simultáneamente, el profesorado emitía su propia evaluación. En todo momento, se trató de constituir un proceso que resultase dinámico. Para ello, se limitó tanto el número de aspectos evaluables como el tiempo de deliberación. Finalmente, las calificaciones emitidas por los alumnos fueron procesadas por el profesorado.

Debido al número total de trabajos realizados, las exposiciones se han dividido en dos sesiones de 120 minutos de duración (turnos: semana 1 y semana 2). En este sentido, dado que la mitad de los alumnos ya habría actuado como evaluadores con anterioridad a su exposición (aquellos que habrían expuesto durante el turno semana 2 y evaluado a los grupos del turno semana 1), se ha realizado una comparativa de los resultados obtenidos en cada turno.

A continuación, se detalla el procedimiento de aplicación, procesado de las calificaciones y evaluación del proceso utilizado.

3.1. Aplicación combinada de la autoevaluación y evaluación por pares.

La incorporación del alumnado a la evaluación se ha realizado durante tres cursos académicos (20/21, 21/22 y 22/23). En la Tabla 2 se recogen los datos de la muestra y la tipología de la acción desarrollada.

Tabla 2. Diseño del sistema de evaluación con participación del alumnado.

Curso	Alumnos	Grupos (N)	AE^a	EP^a	P^a	Descripción de la acción
20/21	27	10	No	Si	Si	Evaluación con participación del alumnado (P y EP)
21/22	30	10	Si	Si	Si	Evaluación combinada (P y EP/AE)
22/23	33	11	Si	Si	Si	Evaluación combinada (P y EP/AE) y seguimiento

^a AE: Autoevaluación; EP: Evaluación por pares; P: Evaluación del profesorado.

En este sentido, cabe señalar que durante la primera aplicación de la innovación (curso 20/21) se empleó únicamente la evaluación por pares (EP) sin referencia alguna a la autoevaluación (AE). No obstante, debido a que los resultados sugerían un cierto sesgo y/o desviación de una parte de las calificaciones

emitidas, en los dos siguientes cursos (21/22 y 22/23) la evaluación por pares se planteó de un modo combinado con la autoevaluación.

En este sentido, se estableció como criterio que la evaluación por pares estuviese, además, referenciada a la autoevaluación. De este modo, cada alumno debía emitir una calificación del resto de trabajos de un modo comparativo con su propio trabajo. El rango de la variación, o calificación comparativa (CC), se limitó al intervalo [-2.5, +2.5] puntos. Para facilitar la comprensión del procedimiento se estableció un valor de referencia de 7.5 puntos para el trabajo propio. De este modo, los alumnos evaluadores podían emitir una calificación dentro del intervalo [5, 10] puntos, con el que, mayoritariamente, se sentían más familiarizados. Así, por ejemplo, una calificación emitida de 9 puntos equivalía a un valor de CC de +1.5 puntos, lo que conllevaba una mejor valoración del trabajo evaluado en comparación con el propio.

Para la obtención de la calificación por pares correspondiente a cada uno de los trabajos (EP_i), se promedió el conjunto de calificaciones recibidas por dicho trabajo del conjunto de evaluadores (resto de alumnos).

La calificación de autoevaluación de cada grupo (AE_i) se obtuvo mediante la diferencia entre el valor de EP_i y la calificación comparativa (CC_i) emitida por los componentes del grupo autoevaluado para dicho trabajo i . El valor final de la calificación se obtuvo del promedio de las diferencias ($EP_i - CC_i$) obtenidas para los N grupos evaluados, tal y como se muestra en la ecuación 1.

$$AE_i = \sum_{i=1}^N (EP_i - CC_i) / N \quad (1)$$

3.2. Estandarización de las calificaciones de los pares y profesorado.

Falchikov y Goldfinch (2000) señalan que los temores del profesorado acerca de la falta de confiabilidad o validez de la evaluación por pares pueden restringir su uso privando, de este modo, a muchos estudiantes del beneficio que este método conlleva para su aprendizaje.

Con el objetivo de reducir las desviaciones entre las calificaciones del profesor (P_i) y los pares (EP_i) realizadas sobre un mismo trabajo, se ha propuesto un procedimiento de estandarización de dichas calificaciones. Para ello, se ha considerado, para el conjunto de trabajos calificados (N), la calificación promedio correspondiente a las evaluaciones del profesor (P_{med}) y de los pares (EP_{med}).

A continuación, ponderando los promedios anteriores, se ha estimado la calificación promedio global (C_{med}), tal y como indica la ecuación 2, donde el parámetro p representa el peso correspondiente a la evaluación del profesorado y puede presentar valores en el intervalo [0,1]. De este modo, la presencia de p otorga cierta capacidad al profesor sobre el control del proceso, permitiéndole definir el peso de cada conjunto. No obstante, en este estudio se ha considerado en todo momento un valor de $p=0.5$, otorgando así el mismo peso a la evaluación del profesor y de los pares.

$$C_{med} = p \cdot P_{med} + (1 - p) \cdot EP_{med} \quad (2)$$

Por último, los valores de la calificación estandarizada correspondiente a cada trabajo i evaluado ($P_{i,est}$ y $EP_{i,est}$) se obtuvieron por corrección de sus respectivas calificaciones originales (P_i y EP_i) mediante el uso de las ecuaciones 3 y 4.

$$EP_{i,est} = EP_i \cdot \frac{C_{med}}{EP_{med}} \quad (3)$$

$$P_{i,est} = P_i \cdot \frac{C_{med}}{P_{med}} \quad (4)$$

3.3. Seguimiento y evaluación de la percepción del alumnado.

Para medir las percepciones de los estudiantes sobre el proceso de evaluación por pares, tras la finalización del curso, se realizó al alumnado participante la encuesta mostrada en la Tabla 3. La encuesta recoge 12 preguntas, divididas en cinco secciones referentes a su percepción acerca del desarrollo del método, sobre su objetividad, su participación como evaluador o evaluado y su inclusión en el sistema de evaluación. Para la evaluación de las opiniones se utilizó una escala de cuatro puntos correspondientes a las categorías: totalmente en desacuerdo (1), parcialmente en desacuerdo (2), parcialmente de acuerdo (3) y totalmente de acuerdo (4).

Tabla 3. Encuesta al alumnado sobre su percepción acerca de la autoevaluación y evaluación por pares.

Criterio	Cuestión (Q)
Desarrollo	1. Los estudiantes no deben involucrarse en el proceso de evaluación, sólo debe evaluar el profesor.
	2. Participar en la evaluación por pares me requiere mucho tiempo y esfuerzo.
	3. Los criterios de evaluación están preestablecidos y son claros.
Objetividad	4. Cuando evalúo a mis compañeros siempre soy justo y objetivo.
	5. La relación personal con mis compañeros influye cuando califico su trabajo.
Rol evaluado	6. Considero que mis compañeros están cualificados para evaluar mi trabajo.
	7. La retroalimentación recibida cuando me evalúan mis compañeros me ayudan a mejorar mi trabajo.
Rol evaluador	8. Evaluar a mis compañeros consigue que esté más concentrado durante sus exposiciones.
	9. Evaluar a otros compañeros me ayuda a reflexionar sobre mi propio trabajo y mejorarlo.
	10. Comparar otros trabajos con el mío me ayuda a desarrollar un pensamiento crítico.
Evaluación	11. En la calificación del trabajo, el peso de la evaluación por pares debería ser inferior al peso del profesor.
	12. Recomendaría que la evaluación por pares formase parte del sistema de evaluación.
Valoración	Categoría
1	Totalmente en desacuerdo (TD)
2	Parcialmente en desacuerdo (PD)
3	Parcialmente de acuerdo (PA)
4	Totalmente de acuerdo (TA)

4. Resultados.

Los resultados de las calificaciones otorgadas por cada evaluador: profesor (P), pares (EP) y autoevaluación estimada (AE), junto con las calificaciones estandarizadas y la desviación entre los grupos profesor-pares y pares-autoevaluación se muestran en la Tabla 4.

En primer lugar, cabe señalar que se ha observado una evolución de la desviación entre la calificación del profesor y los pares (P-EP). En este sentido, durante el curso 20/21 todas las calificaciones procedentes del alumnado superaron a la del profesor con desviaciones entre [-0.36, -1.47] puntos. En este curso, el proceso se centró en la calificación de los trabajos de los pares basándose en los criterios establecidos. Sin embargo, la tendencia se modifica durante los siguientes cursos en los que la valoración se estableció de modo comparativo al trabajo propio del evaluador. En este sentido, el curso 21/22 la desviación sigue mostrando una mayor calificación por parte de los pares aunque con una desviación inferior, entre [-0.95, 0.01] puntos, mientras que en el curso 22/23 la tendencia es invierte, siendo inferiores las calificaciones otorgadas por los pares, con desviaciones entre [0.05, 1.41] puntos. Este cambio en la tendencia coincide con, por una parte, una variación en la calificación del profesor y, por otra parte, con una contención de las calificaciones

otorgadas por los pares. Así, mientras en el curso 20/21 el 80% de las calificaciones de los pares emitidas superaba el umbral de los 8 puntos (40% superior a 9 puntos), en los cursos 21/22 y 22/23 sólo un 53% superaba los 8 puntos (26% y 13% superior a 9 puntos, respectivamente). Probablemente, este efecto puede estar asociado a cierta contención del efecto emocional, junto con un incremento del proceso reflexivo, debido a tener que valorar el trabajo de los compañeros asociado a la diferencia con respecto a su propio trabajo.

Tabla 4. Calificaciones de la evaluación participativa, calificaciones estandarizadas y desviación entre grupos.

Curso	Calificación	P	EP	AE	P _{est} ^a	EP _{est} ^a	Nota ^b	P-EP ^c	P _{est} -EP _{est} ^c	EP-AE ^c
20/21	Media	7.75	8.64		8.20	8.20	8.20	-0.89	0.00	
	DS	0.54	0.29		0.57	0.28	0.41	0.34	0.38	
	Mínima	6.75	8.17		7.14	7.75	7.44	-1.47	-0.61	
	Máxima	8.75	9.11		9.25	8.64	8.95	-0.36	0.61	
21/22	Media	7.39	7.86	7.86	7.62	7.62	7.62	-0.47	0.00	0.00
	DS	0.87	0.77	0.43	0.90	0.74	0.81	0.32	0.34	0.79
	Mínima	5.88	6.60	6.90	6.06	6.40	6.23	-0.95	-0.50	-0.94
	Máxima	8.50	8.97	8.35	8.77	8.70	8.74	0.01	0.54	1.44
22/23	Media	8.44	7.86	7.83	8.15	8.15	8.15	0.58	0.00	-0.02
	DS	0.84	0.62	0.56	0.82	0.65	0.72	0.37	0.34	0.72
	Mínima	7.00	6.43	6.67	6.76	6.67	6.72	0.05	-0.46	-1.28
	Máxima	9.60	8.51	8.85	9.27	8.83	8.88	1.41	0.78	1.27

^a Calificación estandarizada del profesorado (P_{est}) y pares (EP_{est}).

^b Calificación final de la actividad ($Nota = 0.5 \cdot P_{est} + 0.5 \cdot EP_{est}$).

^c Desviación entre calificación de pares con profesor ($P-EP$) y autoevaluación calculada ($EP-AE$).

En segundo lugar, la desviación entre la calificación de los pares y el resultado de la autoevaluación estimada (a partir de las desviaciones emitidas por cada evaluador) muestra un rango de variación ligeramente más amplio, situándose para los cursos 21/22 y 22/23, respectivamente, entre [-0.94, 1.44] y [-0.28, 1.27] puntos. A pesar de ello, asociado al método de estimación, los valores promedio se sitúan en mismo rango con una desviación promedio, prácticamente, nula. No obstante, la desviación promedio se sitúa en 0.59 (21/22) y 0.56 (22/23) puntos, al considerar un valor absoluto de las desviaciones calculadas.

Por otra parte, el algoritmo de corrección aplicado a las calificaciones del profesor y de los pares consigue anular la desviación promedio entre dichos conjuntos, manteniendo la amplitud del rango de desviación (20/21: [-0.61, 0.61], 21/22: [-0.50, 0.54] y 22/23: [-0.46, 0.78]) en relación a las calificaciones originales.

A continuación, en los siguientes apartados se realiza un análisis más detallado de los resultados obtenidos. Para ello, se han subdividido éstos en cuatro secciones. En primer lugar, se analiza con mayor detalle la correlación existente entre la evaluación por pares con respecto a la autoevaluación estimada y la calificación del profesor. Por otra parte, se ha analizado el posible impacto inmediato en el aprendizaje y su rendimiento académico. Por último, se discute los resultados relativos a la percepción del alumnado participante en el proceso de evaluación.

4.1. Evaluación por pares y autoevaluación combinada.

En la Fig. 1 se muestra la relación entre la calificación de autoevaluación calculada (ecuación 1) con la calificación obtenida mediante la evaluación por pares. Si bien sólo el 19% de los trabajos presentaban una

desviación (EP-AE) superior a un punto, no se observa una correlación entre ambas calificaciones (curso 21/22: $r^2 = 0.046$ y curso 22/23: $r^2 = 0.067$).

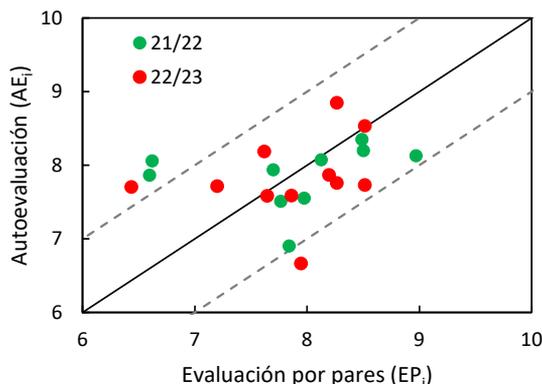


Fig. 1. Relación entre la calificación por autoevaluación calculada (AE_i) y por evaluación por pares (EP_i).

La escala de calificación comparativa (CC_i) utilizada, relativa al trabajo propio en el rango $[-2.5, 2.5]$, permite establecer una valoración relativa al propio para cada uno de los trabajos. En este sentido, a nivel global, se ha observado una ligera tendencia a la estimación positiva del trabajo realizado por los pares frente al trabajo realizado por el evaluador.

La Fig. 2 muestra el histograma de las calificaciones de autoevaluación estimada. Al analizar estos resultados, se puede observar una alta concentración de calificaciones en un intervalo bastante reducido (curso 21/22: 90% entre $[7.5-8.5]$; curso 22/23: 60% entre $[7.5-8.0]$), correspondientes a valores de CC_i en el rango positivo $[0.0-0.5]$ puntos. Este efecto, junto con la falta de correlación observada, sugiere que el valor estimado para la autoevaluación no debería ser incluido con fines sumativos en la calificación final.

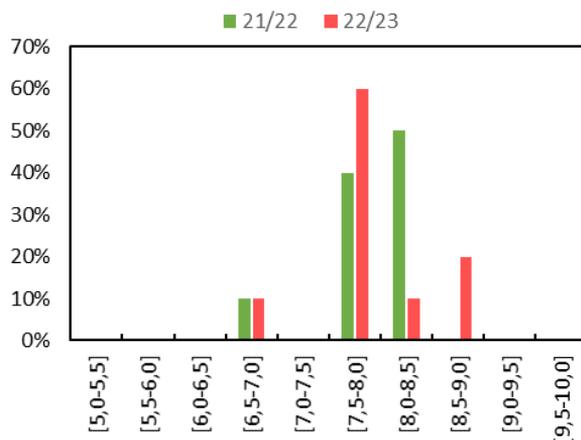


Fig. 2. Histograma de calificaciones correspondientes a la autoevaluación comparativa calculada (AE_i).

Sin embargo, consideramos que resulta beneficioso mantener dicho sistema de calificación comparativa ya que, a efectos formativos, sí que implica un proceso de autorreflexión que comporta ventajas evidentes para su aprendizaje. Entre estas ventajas, cabe señalar que la comparación reflexiva ayuda al estudiante a tomar conciencia de la calidad de su propio trabajo, contribuyendo así a su mejora, y ello les ayuda a desarrollar

la objetividad requerida para alcanzar una autoevaluación efectiva. De hecho, tal y como señala Brooks (2004), la evaluación formativa sólo alcanzaría su máximo potencial cuando los alumnos están involucrados a través de su autoevaluación.

4.2. Correlación entre la evaluación por pares y la evaluación del profesor.

La posible falta de objetividad, junto con la falta de confianza en la imparcialidad del alumnado en su rol de evaluador de sus pares, generan en el profesorado un cierto temor de falta de fiabilidad de la evaluación por la participación del alumnado en el proceso. En este sentido, se ha analizado la correlación existente entre las calificaciones procedentes de los pares (EP_i) y del profesor (P_i). Tal y como puede observarse en la Fig. 3a sí que existe una correlación positiva entre las calificaciones entre EP_i y P_i , con valores del coeficiente r^2 de 0,682 (curso 20/21), 0,871 (curso 21/22), y 0,843 (curso 22/23).

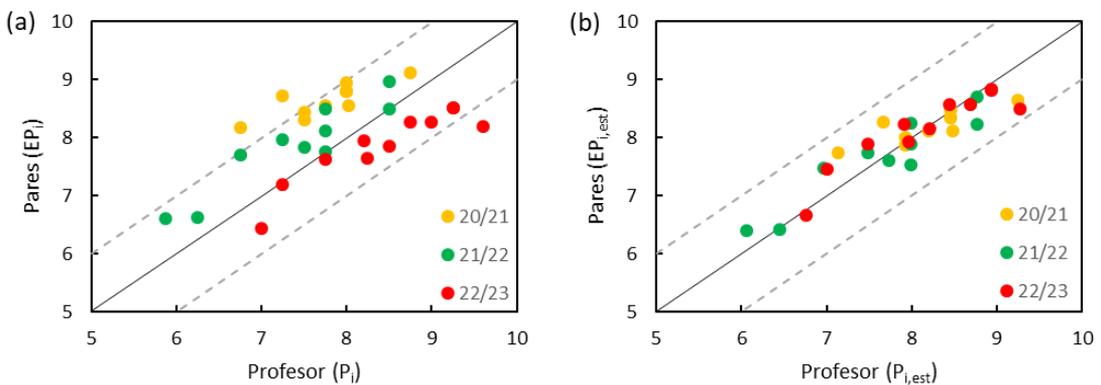


Fig. 3. Relación entre la calificación de los pares (EP) y del profesor (P). (a) Calificación original y (b) estandarizada.

Sin embargo, a pesar de la bondad de la correlación, puede observarse una cierta desviación del conjunto de calificaciones. En este sentido, esta tendencia es más notable para el curso 20/21, en el que la evaluación por pares no implicaba una comparación con el trabajo propio. Por ello, con el objetivo de minimizar dichas desviaciones se propuso una estandarización de las calificaciones (descrita en el apartado 3.2) que condujese a una reducción de la desviación entre dichos grupos y, de ese modo, se generase una mayor confianza en la percepción de su fiabilidad, tanto por parte del profesorado como del alumnado.

De este modo, en la Fig. 3b se muestra la correlación observada entre las calificaciones estandarizadas ($EP_{i,est}$ y $P_{i,est}$) según las ecuaciones 3 y 4. Puede observarse que, si bien se mantiene una misma diferencia relativa entre ellas (con un mismo coeficiente r^2), la pendiente de su ajuste lineal presenta un valor de 0,997 en todos los cursos (equiparable a la recta $EP_{i,est} = P_{i,est}$). Por ello, pensamos que esta mejor correspondencia puede ser percibida positivamente por el profesorado, a efectos de fomentar el uso de estos sistemas de evaluación, siendo incluidos en el sistema de evaluación con un carácter sumativo.

A pesar de que este proceso requiere un mayor esfuerzo de procesado de la información, consideramos que deberían prevalecer otros beneficios existentes en la decisión de incluir la evaluación por pares como un componente del sistema de evaluación. Entre otras consideraciones, cabe destacar que la incorporación al sistema de evaluación aporta variedad en la metodología de evaluación y aprendizaje, constituye una oportunidad para monitorizar la evolución del alumnado y mejora la práctica profesional como docentes.

4.3. Impacto en el rendimiento académico.

La incidencia de la participación del alumnado en la evaluación sobre su rendimiento académico inmediato ha sido otro de los aspectos que se ha analizado. En este sentido, se ha clasificado al alumnado en dos grupos, según la incidencia de su proceso de autorreflexión en su tarea de exposición oral. Así, mientras los alumnos del *turno semana 1*, todavía no habían actuado como evaluadores con antelación a su propia exposición (o lo habían hecho durante la misma sesión), los alumnos del *turno semana 2* habrían actuado con antelación como evaluadores. De este modo, los alumnos del *turno semana 2* habrían experimentado, al evaluar a sus pares, un proceso autorreflexivo sobre su propio trabajo previo a su presentación.

En la Fig. 4 se ha muestran los resultados de la diferencia entre la calificación promedio obtenido por ambos grupos (*turno semana 2 y 1*), tanto para las calificaciones otorgadas tanto por el profesor (a) como por los pares (b). En este sentido, se ha medido la diferencia en la calificación otorgada, tanto en aspectos relativos al contenido (dominio del tema, selección de contenidos, orden y coherencia) como en aspectos relativos a la exposición oral (comunicación y lenguaje, uso y calidad de materiales audio-visuales, adecuación de la duración y puesta en escena).

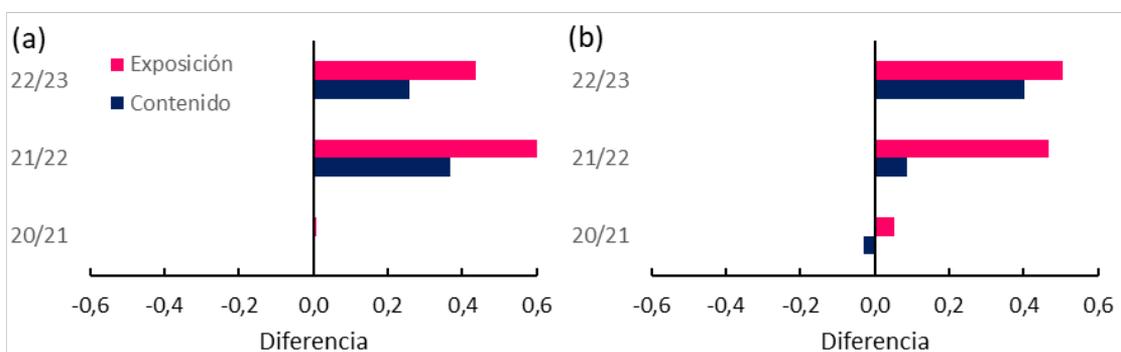


Fig. 4. Diferencia entre la calificación promedio de los alumnos del turno número 2 (con participación previa como evaluadores) y turno número 1 (sin participación previa como evaluadores). (a) Evaluación del profesor y (b) evaluación por pares.

Si bien el conjunto de trabajos analizados es reducido, los resultados obtenidos durante los cursos 21/22 y 22/23 muestran que existe, tanto para la calificación del profesor como de los pares, cierta ventaja en la valoración global de los trabajos del turno 2, respecto a los del turno 1. Esta mejora, aproximadamente, se sitúa en 0,25 puntos en relación a la valoración del contenido y en 0,50 puntos en relación a la valoración de la exposición oral. Si bien estas diferencias no son excesivas, y considerando que también podrían estar asociadas con otros factores, no deja de ser llamativo que, únicamente, se hayan observado en los cursos en los que se empleó una calificación comparativa. Así, en el curso 20/21, que no se incluyó dicha referencia al trabajo propio, no se apreció diferencias entre ambos turnos.

Esta mejora podría ser entendida, en parte, por alcanzarse una mejor comprensión de los diferentes niveles de calidad asociados a la actividad objeto de evaluación. Reinhorlz (2016) señala que la evaluación por pares mejora la capacidad de autoevaluación al explicitar la gradación de los estándares de calidad. Por ello, cuando un alumno actúa como evaluador de sus pares, de modo autorreflexivo, su trabajo mejora.

Por otra parte, durante las exposiciones públicas, se ha observado un comportamiento más activo por parte del alumnado asistente. En este sentido, entre las ventajas de su aplicación de la evaluación por pares, cabe considerar que se promueve la interacción y la discusión con otros compañeros durante la exposición (por ejemplo, compartiéndose observaciones entre evaluadores) y tras su finalización (efectuándose una retroalimentación formativa). Además, en el ámbito del desarrollo de habilidades y competencias

transversales, existen también beneficios de carácter formativo, como la promoción de la comunicación efectiva, el aprendizaje autorreflexivo y el pensamiento crítico.

4.4. Percepción del alumnado.

La encuesta mostrada en la Tabla 3 se ha diseñado para obtener la percepción del alumnado acerca del proceso de evaluación por pares implementado. En este sentido, se ha evaluado su percepción sobre el desarrollo, la confianza, su percepción siendo evaluador y/o evaluado por sus pares y, por último, su inclusión como parte de la evaluación sumativa. Para ello, se ha utilizado una escala de Likert de cuatro puntos, siendo éstos agrupados en dos bloques principales (acuerdo/desacuerdo). A pesar que este seguimiento de la percepción del alumnado únicamente se ha llevado a cabo en el curso académico 22/23 con una muestra de 22 respuestas, se ha tratado de realizar un análisis de los resultados obtenidos, mostrados en la Fig. 5.

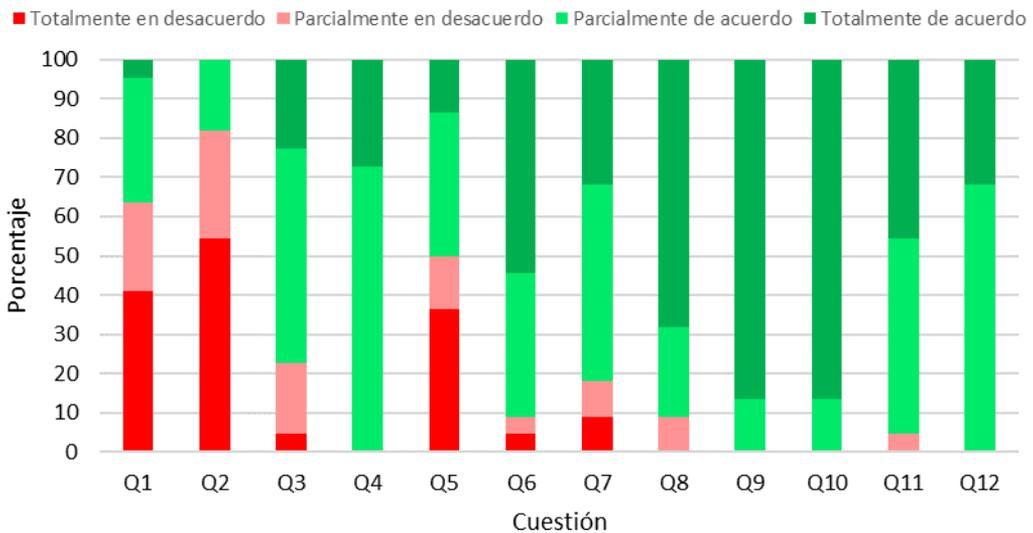


Fig. 5. Opinión de los estudiantes sobre el proceso de evaluación por pares realizado durante el curso 22/23 (n=22).

En primer lugar, en relación a la aplicación de la evaluación por pares, puede resultar sorprendente que un conjunto elevado de alumnos considere que la evaluación únicamente debería involucrar al profesorado (Q1: 36,4%). Por otra parte, en referencia a desarrollo, mayoritariamente, los alumnos consideran que el proceso no les requiere demasiado tiempo y/o esfuerzo (Q2: 81,8%) y que los criterios están preestablecidos y son claros (Q3: 77,3%).

En segundo lugar, en referencia a su percepción acerca de la confiabilidad del proceso, aunque el 100% de los alumnos coinciden en considerar que actúan de forma honesta al evaluar a sus compañeros (Q4: un 27% lo está completamente), el 50% de las respuestas (Q5) consideran que la relación personal con sus compañeros influye durante la evaluación a sus compañeros. Este tipo de temor o percepción negativa acerca de la fiabilidad de los resultados está comúnmente presente en este tipo de evaluación participativa. En concreto, revelaron que aquellos alumnos con un menor nivel de autoconocimiento reaccionan más negativamente al recibir la evaluación por pares (Carlsson y Nilsson, 2022).

En tercer lugar, en relación a su percepción de la capacidad de sus compañeros para actuar como evaluadores, cabe considerar que los alumnos encuestados se encuentran en el cuarto curso y la actividad

objeto de evaluación no involucra aspectos demasiado técnicos y/o complejos. Con ello, mayoritariamente, los alumnos coinciden en considerar que sus compañeros están capacitados para actuar como evaluadores (Q6: 91.9%) y, además, consideran que la retroalimentación recibida, al ser evaluados, les resulta de ayuda para mejorar en su trabajo (Q7: 81.8%).

En cuarto lugar, en referencia a su experiencia como evaluador de sus compañeros, los alumnos perciben, mayoritariamente, que el proceso logra mejorar su nivel de atención (Q8: 90,9%) y consideran que contribuye a poder reflexionar sobre su propio trabajo y el margen de mejora existente, así como a desarrollar un pensamiento crítico durante el proceso (Q9 y Q10: 100%, 86,4% completamente). En este sentido, cabe destacar la percepción positiva que el alumnado muestra ante el impacto que el proceso conlleva en su aprendizaje y desarrollo de habilidades.

Por último, en referencia a su repercusión en el sistema de evaluación de la asignatura, si bien, los alumnos consideran que el peso de la evaluación por pares no debería equipararse al del profesorado, debiendo predominar este último (Q11: 95,5%), coinciden en la conveniencia de ser incluido en el sistema de evaluación de la asignatura (Q12: 100%).

No obstante, debe tenerse en cuenta que el presente estudio se ha realizado en alumnos de cuarto curso de grado, cuya percepción sobre el valor de la retroalimentación recibida por sus compañeros puede diferir significativamente de aquellos estudiantes que terminen de acceder a la educación universitaria, que pueden tener percepciones más negativas y mayores reservas para comprometerse con estas prácticas (Poverjuc et al., 2012).

5. Conclusiones.

A partir de la experiencia docente desarrollada, puede deducirse que, aunque todavía existen ciertas barreras que tienden a evitar la inclusión de la participación del alumnado en la evaluación, los beneficios que conlleva deberían aconsejar su consideración como parte del proceso de evaluación formativa.

Según indica Boud (1999), la principal preocupación debería perseguir que las tareas de evaluación resulten favorables al aprendizaje entre iguales, mediante la promoción de prácticas positivas y evitando aquellas que lo inhiben. En este sentido, entre las limitaciones existentes, cabe incluir la consideración del profesorado de que existe una respuesta más emocional que objetiva por parte del alumno evaluador, o su posible falta de comprensión o inadecuada interpretación de los criterios establecidos en la evaluación, así como cierta subjetividad y dificultad de aplicación para determinados contextos o actividades, sin olvidar el mayor esfuerzo que se requiere, por parte del profesor, tanto en su diseño y planificación, como en la etapa de procesado. Entre los principales logros de carácter formativo que suponen para el aprendizaje del alumnado se incluye el aumento en la toma de consciencia sobre su proceso de aprendizaje, el aprendizaje motivado por la visión de sus compañeros, el fomento de la interacción entre el alumnado, el aumento de confianza en el sistema de evaluación.

Para fomentar la participación del alumnado en el proceso de evaluación se requiere una planificación previa que considere una selección de las actividades más adecuadas, así como el establecimiento claro de las reglas y criterios de evaluación, junto con el uso de una estrategia de aplicación y desarrollo que no resulte demasiado rígida o requiera un esfuerzo excesivo que pudiese constituir una carga adicional al profesorado que conllevara una limitación en su uso (Marin-García et al., 2008).

El uso de acción combinada de calificación de los pares con la comparativa hacia el trabajo propio conlleva un proceso de autorreflexión que disminuye el riesgo de una respuesta más emocional y potencia la confianza en las calificaciones resultantes. No obstante, si bien el valor de la autoevaluación diferido (a partir de las calificaciones comparativas) puede resultar inadecuado a efectos de poder ser considerado como parte de la calificación final de la actividad, sí que contribuye favorablemente en el proceso de aprendizaje, dado que ayuda al alumno a desarrollar una mayor objetividad y capacidad de autorreflexión.

Por el contrario, se recomienda la incorporación al sistema de evaluación de las calificaciones procedentes de la evaluación por pares, dado que se ha observado una adecuada correlación con las emitidas por el profesorado. En la misma línea, Marin-García et al. (2008) consideran que incluir las calificaciones de los pares en la estimación de la evaluación final no supondría una distorsión de los resultados. Esta incorporación se justificaría, no sólo por su falta de distorsión de los resultados, sino porque ello incrementa el grado de responsabilidad adquirido por los estudiantes al actuar como evaluador. Además, el empleo de un algoritmo de estandarización de las calificaciones de pares y profesor consigue hacer comparables ambas calificaciones, reduciéndose así la percepción de falta de confianza en su fiabilidad.

En relación a la actuación del alumnado como evaluador, se ha observado que ésta ha contribuido, con carácter inmediato, a una ligera mejora de su rendimiento académico. Probablemente, ello puede estar asociado al desarrollo de su capacidad de autorreflexión. Además, se ha observado un desarrollo de sus habilidades y competencias de carácter transversal, especialmente, la comunicación efectiva, el pensamiento crítico y el aprendizaje autorreflexivo.

Por último, los resultados de la encuesta de seguimiento muestran que una parte significativa del alumnado considera que el proceso de evaluación no les compete. Además, señalan que consideran ser justos mientras actúan como evaluadores, aunque la mitad ellos consideran que la relación personal influye en su valoración. En este sentido, su acción como evaluadores contribuye a mejorar su nivel de atención al resto de trabajos, a desarrollar un pensamiento crítico y a reflexionar sobre su propio trabajo. Del mismo modo, mayoritariamente, consideran que sus compañeros están capacitados para evaluarles y sus retroalimentaciones les resultan de ayuda. Finalmente, consideran que la evaluación por pares debería estar incluida en el sistema de evaluación de la asignatura, pero con un peso inferior a la evaluación del profesor.

6. Referencias.

- Atarés Huerta, L., Llorens Molina, J. A. y Marin-García, J.A. (2021). La evaluación por pares en Educación Superior. *Educación Química*, 32(1), 112-121.
- Boud, D., Cohen, R. y Sampson, J. (1999). Peer learning and assessment. *Assessment & evaluation in higher education*, 24(4), 413-426.
- Carlsson Hauff, J. y Nilsson, J. (2022). Students' experience of making and receiving peer assessment: the effect of self-assessed knowledge and trust. *Assessment & Evaluation in Higher Education*, 47:6, 959-971. <https://doi.org/10.1080/02602938.2021.1970713>
- Dochy, F. y Moerkerke, G. (1997). The present, the past and the future of achievement testing and performance assessment. *International Journal of Educational Research*, 27(5), 415 - 432
- Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self assessments. *Assessment and Evaluation in Higher Education*, 11(2), 146-166.

- Falchikov, N. y 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.
- Ibarra, M., Rodríguez, G. y Gómez, R. (2012). La evaluación entre iguales: beneficios y estrategias para su práctica en la universidad. *Revista de Educación*, 359(2), 1-19.
- Magin, D. y Helmore, P. (2001). Peer and teacher assessments of oral presentation skills: how reliable are they? *Studies in Higher Education*, 26(3), 288-297.
- Marin-Garcia, J.A., Miralles Insa, C. y Marin Garcia, P. (2008). Oral Presentation and Assessment Skills in Engineering Education. *International Journal of Engineering Education*, 24(5), 926-935.
- Poverjuc, O., Brooks, V. y Wray, D. (2012). Using peer feedback in a Master's programme: a multiple case study. *Teaching in Higher Education*, 17(4), 465-477.
- Reinholz, D. (2016). The assessment cycle: A model for learning through peer assessment. *Assessment & Evaluation in Higher Education*, 41(2), 301-315.
- Romá, M., Ballester-Berman, J.D., Vives, F.P., López-Sánchez, J.M., Signes, J.A., Martínez-Marín, T., Martín, E. y Selva, J. (2022, July 6). Strengths and dangers of self and peer assessment in engineering learning. Teachers' and students' perspective. International Symposium on Project Approaches in Engineering Education; Active Learning in Engineering Education Workshop; International Conference on Active Learning in Engineering Education (PAEE/ALE'2022), Alicante - Spain. <https://doi.org/10.5281/zenodo.6924152>
- White, E. (2009). Student perspectives of peer assessment for learning in a public speaking course. *Asian EFL Journal*, 33(1), 1-36.