

# Autoevaluación, co-evaluación y estudio empírico frente a técnicas de evaluación tradicionales en la asignatura de Sistemas Operativos

Patricia Martín-Rodilla, Javier Parapar  
Information Retrieval Lab  
University of A Coruña  
A Coruña, Spain

[patricia.martin.rodilla@udc.es](mailto:patricia.martin.rodilla@udc.es), [javierparapar@udc.es](mailto:javierparapar@udc.es)

## Resumen

En este artículo, presentamos un protocolo desarrollado como parte de las actividades de innovación docente en Sistemas Operativos. El objetivo del protocolo es incorporar la autoevaluación como mecanismo de evaluación formativa en competencias en Sistemas Operativos, que fomente en el estudiante capacidades transversales de especial relevancia en diseño e implementación de funcionalidades de Sistemas Operativos: análisis crítico, detección de posibles mejoras y conciencia del propio proceso de aprendizaje. Para ello, se diseñó un protocolo de autoevaluación y co-evaluación de las pruebas prácticas de la asignatura correspondientes a tres áreas temáticas en Sistemas Operativos: sistemas de ficheros, manejo de memoria y gestión y planificación de procesos. Buscando posibilitar la convivencia de ambos sistemas de evaluación, el protocolo se aplicó a parte del alumnado, manteniendo la evaluación por parte del profesorado paralelamente. Posteriormente, se realizó un estudio empírico de precisión en auto-evaluación y co-evaluación de los estudiantes, comparando el protocolo con la evaluación tradicional, y evaluando inicialmente sus implicaciones en las calificaciones finales obtenidas. Los resultados permiten no sólo evaluar inicialmente el protocolo diseñado, sino también conocer el punto de partida en capacidades de autoevaluación y co-evaluación de los estudiantes en las áreas temáticas particulares de Sistemas Operativos.

## Abstract

In this paper, we present a protocol developed as part of some educational innovation initiatives in Operating Systems subject. The protocol's goal is incorporating self-evaluation as a formative evaluation mechanism in Operating Systems that fosters in university students of computer science the skills of

critical analysis, detection of possible improvements and awareness of their own learning process, all of them of special relevance in the design and implementation of functionalities of Operating Systems. For that purpose, a self-evaluation and co-evaluation protocol was designed for the practical exercises corresponding to three thematic areas in Operating Systems: file systems, memory management, and process management and scheduling. In order to facilitate the coexistence of both evaluation systems, the protocol was applied to part of the students, maintaining the traditional evaluation by the teaching staff in parallel with all the students. Subsequently, an empirical study of precision in self-evaluation and co-evaluation of the students was carried out, analyzing the correlation with the traditional evaluation, and the implications of the designed protocol in the final grades obtained. The results allow not only to initially evaluate the designed protocol, but also to know the starting point in self-evaluation and co-evaluation capacities of the students in Operating Systems areas.

## Palabras clave

Autoevaluación, co-evaluación, Sistemas Operativos, estudio empírico.

## 1. Contexto docente e investigador

La docencia en Ingeniería Informática ha sufrido una profunda transformación en los últimos años, impulsada por los cambios de paradigma como la adaptación al EEES [3, 8, 18] o, recientemente, a los espacios virtuales de enseñanza y aprendizaje. En este contexto, las asignaturas de los primeros cursos en Ingeniería Informática tienen por objetivos principales la adquisición y desarrollo de las capacidades y competencias transversales en el alumno. Estudios previos han demostrado mejora en este tipo de capacidades a través de la aplicación de innovaciones en

los mecanismos de evaluación [6]. Tras la detección de problemas asociados a capacidades transversales en nuestra asignatura de Sistemas Operativos, este artículo presenta un protocolo de autoevaluación y co-evaluación (evaluación por pares) de las pruebas prácticas de Sistemas Operativos, así como un estudio empírico de resultados del primer año de implantación de dicho protocolo. El objetivo es emplear la autoevaluación y la co-evaluación como mecanismo que fomente en los alumnos las capacidades de análisis crítico, detección de posibles mejoras y conciencia de su propio proceso de aprendizaje, capacidades todas ellas de especial relevancia en diseño e implementación de funcionalidades y estructuras de Sistemas Operativos.

A continuación, se realiza un recorrido por las iniciativas de innovación docente en Sistemas Operativos puestas en marcha en los últimos años en el contexto educativo español, así como otros trabajos y experiencias de autoevaluación y co-evaluación en contextos de Ingeniería Informática que ayudan a situar el protocolo desarrollado.

### 1.1. Innovación docente en Sistemas Operativos

La adaptación de la asignatura de Sistemas Operativos al EEES [4, 8, 11] supuso el inicio de actividades de innovación docente en Sistemas Operativos, comenzando con iniciativas centradas en la propia ejecución de la docencia, y encaminadas a dar más peso al aspecto práctico, asegurar el no abandono de los estudiantes y aumentar su motivación, con resultados satisfactorios [4, 10]. En Canaleta [4] se analizan este tipo de innovaciones docentes en Sistemas Operativos a lo largo de 13 años, siendo el mayor estudio histórico hasta el momento. Posteriormente, las iniciativas de innovación docente en Sistemas Operativos han puesto el foco en otros aspectos, destacando: 1) internacionalización con docencia en otros idiomas (principalmente inglés) o iniciativas bilingües [1], 2) fomento del trabajo continuo y en equipo del alumnado, con técnicas como el *flipped classroom* [24] o la evaluación por pares [9], 3) estudio de ciertas competencias, como la capacidad de organización y planificación del estudio por parte del estudiantado [17] o de análisis visual y seguimiento de la clase [7, 21], con técnicas de mapas conceptuales [20] y 4) adaptación a la práctica de la docencia de Sistemas Operativos en entornos virtuales de aprendizaje o automatización de ciertas tareas de evaluación [13], con técnicas muy diversas [16].

Como puede verse en estos trabajos, las innovaciones docentes en Sistemas Operativos se dan en su mayoría en su adecuación al EEES, y luego a entornos virtuales de aprendizaje. El rango de técnicas aplicadas y experiencias docentes publicadas con este

fin es alto, aunque probadas en plazos cortos de tiempo [4].

Sin embargo, el número de trabajos centrados en fomentar competencias genéricas y capacidades transversales dentro de la asignatura es significativamente menor. En este sentido, la mayor parte de técnicas e iniciativas docentes son aplicadas a aspectos de la propia práctica docente, como *flipped classroom* [24] o mapas conceptuales [20], o estudios estadísticos de resultados a posteriori [13], no encontrando trabajos en Sistemas Operativos que ahonden en la influencia de los cambios en los mecanismos de evaluación como una forma de fomentar dichas capacidades genéricas y/o transversales.

### 1.2. Autoevaluación y co-evaluación en Ingeniería Informática

Autoevaluarse supone implicarse de forma explícita en la revisión del trabajo o actividad propio [6]. Se ha usado el término co-evaluación en estudios en docencia para referirse a varios aspectos que conviene distinguir: 1) a la evaluación por parte de pares, 2) a la evaluación conjunta del trabajo en equipo o 3) a la coexistencia de varios sistemas de evaluación. En este artículo, el protocolo propuesto se refiere a coevaluación como a la evaluación que realizan los alumnos de sus prácticas realizadas en equipo junto con otros compañeros (acepción 2).

Pese a la ausencia de trabajos en docencia en Sistemas Operativos (como se detalla en la sección anterior) que específicamente traten aspectos de evaluación como mecanismo de mejora de capacidades genéricas y/o transversales, son numerosos los estudios y aplicaciones en otras áreas y campos, incluso dentro de los estudios de Ingeniería Informática.

En primer lugar, podemos encontrar trabajos centrados en la aplicación de la autoevaluación y/o la co-evaluación en asignaturas concretas de la Ingeniería Informática, como en asignaturas de programación [14], redes [5] o bases de datos [15], mostrando buenos resultados tanto en la valoración de la experiencia por parte del alumnado, como en resultados académicos [14, 15].

En segundo lugar, estudios de revisión en mayor profundidad sobre estos mecanismos en la docencia de la Ingeniería Informática, señalan también la utilidad de estos mecanismos de autoevaluación para alcanzar una evaluación continua real [14], mejorar la efectividad de los mecanismos de retroalimentación [25], como medio formativo en sí mismo [14, 18] e incluso en términos de mejora del sentido crítico y capacidad de reflexión sobre el propio proceso de aprendizaje [25].

## 2. Objetivos

En este contexto, el objetivo es diseñar un protocolo de autoevaluación y co-evaluación por pares de las prácticas de la asignatura de Sistemas Operativos, como mecanismo de evaluación formativa sobre las competencias específicas de la asignatura que, además, fomentan en el estudiantado capacidades transversales (análisis crítico, detección de posibles mejoras y conciencia del propio proceso de aprendizaje). El protocolo diseñado se implanta en la asignatura de Sistemas Operativos 2020-2021, y se compara con el método tradicional de evaluación por parte del profesor aplicado hasta ahora, permitiendo conocer de forma inicial e intuitiva el punto de partida del alumnado, los efectos del protocolo en la mejora de las capacidades transversales abordadas y su impacto en las calificaciones obtenidas por los estudiantes. No buscamos, por tanto, realizar un estudio empírico estadístico exhaustivo que pruebe la adecuación del protocolo a todos los objetivos (en un solo curso académico no tendríamos muestra suficiente), sino más bien plantear las preguntas de investigación que nos surgen y analizar de manera inicial e intuitiva los datos de este primer curso académico de implantación, para así extraer lecciones aprendidas de la experiencia de innovación docente y desarrollar en el futuro un plan de actuación a largo plazo para la mejora del protocolo propuesto.

## 3. Protocolo y estudio empírico

### 3.1. Hipótesis iniciales y protocolo

De forma similar a otras planificaciones de la asignatura de Sistemas Operativos [4], la evaluación tradicional que se realiza en nuestra asignatura se basa en la revisión y calificación del profesor de dos grandes ítems: examen semestral y tres prácticas de programación en lenguaje C [22]. Las prácticas constituyen un 25% del peso en la calificación final. Cada una de las prácticas se centra en un gran bloque temático de la asignatura: P1 (sistemas de ficheros), P2 (manejo de memoria) y P3 (gestión y planificación de procesos). El objetivo es que los alumnos, organizados en parejas, afiancen los conocimientos teóricos y adquieran las competencias y habilidades prácticas en estas tres áreas. En la Figura 1 puede verse el diagrama del proceso de *evaluación tradicional*: los estudiantes, en grupos reducidos (15-20 personas), se distribuyen en parejas y realizan sesiones de *Focus Group* donde diseñan, codifican, realizan pruebas y resuelven dudas con un profesor en sesiones prácticas. Cada práctica (P1, P2 y P3) tiene una fecha máxima de entrega. Posteriormente, cada pareja de estudiantes acude a una sesión de defensa por práctica donde explican su diseño y programa final.

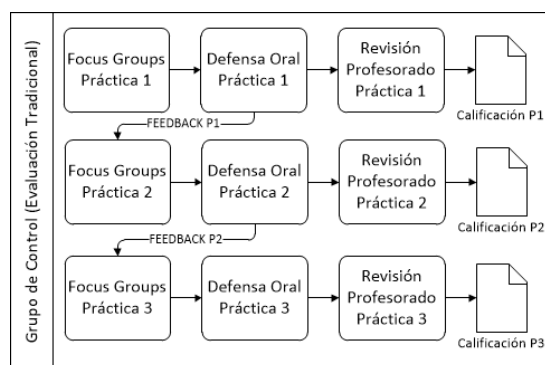


Figura 1: Diagrama de proceso de evaluación tradicional (Grupo de control).

Los estudiantes deben responder a cualquier aspecto sobre la misma para demostrar el conocimiento profundo del trabajo realizado. El profesor califica a la pareja atendiendo a unos criterios estructurales, funcionales y de exposición oral, dando una retroalimentación sobre las funcionalidades incompletas, no entregadas o las dificultades que los estudiantes puedan manifestarle. Aunque los criterios establecidos para cada práctica garantizan la evaluación de forma sistemática, el número de alumnos a evaluar y la escasez de tiempo para cada sesión de defensa no permite dedicar demasiado tiempo a la retroalimentación que el profesor puede ofrecer a cada pareja, siendo esta fundamental para asentar y mejorar nuestras capacidades objetivo (análisis crítico, detección de posibles mejoras y conciencia del propio proceso de aprendizaje). En la Figura 2 puede verse el diagrama del protocolo diseñado.

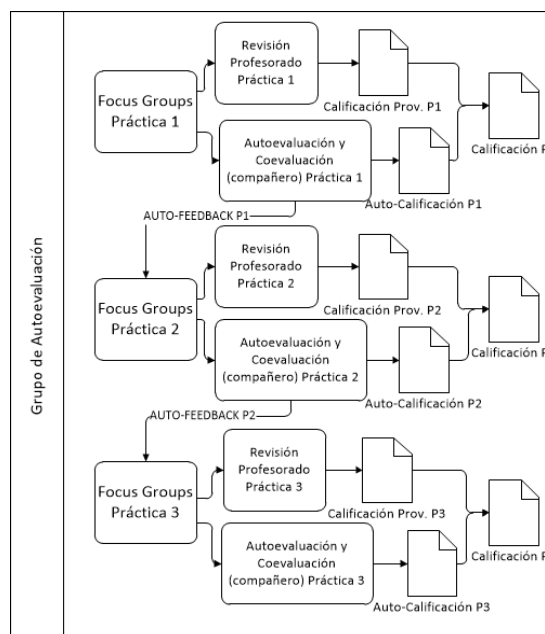


Figura 2: Protocolo de autoevaluación y coevaluación propuesto.

Los estudiantes, al igual que en la evaluación tradicional, atienden a su *Focus Group* con un profesor para el diseño, codificación, prueba y resolución de dudas de cada práctica. Una vez llegada la fecha de entrega de cada práctica, el protocolo establece dos actividades paralelas.

Por un lado, los estudiantes acuden a una sesión de autoevaluación y co-evaluación, donde se les facilita un formulario con una lista de criterios que la práctica a autoevaluar debe cumplir (a modo de *check box*, con posibilidad de anotaciones cortas). Cada pareja debe indicar por cada criterio, si la práctica entregada lo cumple o no y si la funcionalidad relacionada se ejecuta correctamente. Se permiten las anotaciones breves al pie para explicitar, por ejemplo, si detectaron algún error concreto pese a cumplir el criterio completo.

Es importante indicar que los alumnos pueden revisar su código y probar en ejecución su práctica durante la sesión de autoevaluación, así como debatir y revisar el trabajo de su compañero de pareja. Esta parte de coevaluación, pretende que el estudiante encuentre en el compañero/a “un punto de referencia para apreciar si su propia valoración está bien conmensurada” [14], mejorando los resultados de autoevaluación si ésta incluye coevaluación [19].

En paralelo, el profesor realiza una evaluación personal de cada práctica y asigna una calificación provisional “a ciegas”, es decir, sin consultar el formulario de autoevaluación entregado por cada pareja.

Una vez obtenidas sendas calificaciones (Autoevaluación y Provisional del profesor), el docente realiza una ponderación final con ambas como referencia, adecuando la calificación a los criterios reales de cada práctica. Nótese que, siguiendo este protocolo y en comparación con la evaluación tradicional, la retroalimentación obtenida por los estudiantes es diferente, ya que procede: 1) de su propia experiencia mirando el código, ejecutando la práctica y evaluando según los criterios definidos su propio trabajo y 2) de un debate con su compañero acerca de la adecuación de su calificación y trabajo a los criterios definidos, siendo una retroalimentación mucho más detallada que la obtenida con la evaluación tradicional mediante defensa (en tiempo y forma).

### 3.2. Diseño de experimento

Una vez diseñado el protocolo, se decidió realizar un cuasi-experimento, i.e un experimento causa-efecto controlado para entornos donde no es posible seleccionar aleatoriamente los sujetos de la muestra, como el entorno educativo [2].

La elección de los grupos de autoevaluación se realizó de forma aleatoria, si bien su formación viene determinada por las listas de matrícula. La muestra se organiza en grupo de control (evaluación tradicional,

N=75 estudiantes) y grupo de autoevaluación (N=38 estudiantes). La presencia de mayor número de estudiantes en el grupo de control se debe a la ausencia de tiempo por parte del profesorado para aplicar el protocolo de autoevaluación en más grupos, debiéndose balancear en los siguientes cursos académicos.

El objetivo del experimento, siguiendo la metodología de Wohlin [27], se define como: “Comparar el protocolo de autoevaluación y co-evaluación diseñado con el método de evaluación tradicional, con el objetivo de evaluar su influencia en el desarrollo de capacidades genéricas y/ transversales y medir su impacto en las calificaciones y estadísticas finales obtenidas de la asignatura”.

Las preguntas de investigación (RQ) a modo de hipótesis iniciales que definimos son:

*RQ0*: No existen diferencias significativas entre calificaciones del grupo de autoevaluación (las impuestas por los estudiantes, así como las finales) y las calificaciones finales del grupo de control.

Esta hipótesis nos permite analizar implicaciones de capacidad de autoevaluación de los estudiantes.

*RQ1*: No hay diferencia significativa en los cambios en las calificaciones desde P1 a P3 entre ambos grupos.

Esta hipótesis nos permite analizar si es mejor el *feedback* propio (asociado a la capacidad transversal de análisis crítico y detección de mejoras) que el obtenido en defensa por parte del profesor (grupo de control), materializándose en ese caso en que el grupo de autoevaluación mejoraría más en calificaciones de P1 a P3.

*RQ2*: La correlación entre las calificaciones establecidas por los estudiantes en el grupo de autoevaluación y las provisionales del profesor en ese mismo grupo no cambia con el tiempo.

Esta hipótesis nos permite analizar si la autoevaluación ayuda a tomar conciencia del propio proceso de aprendizaje, manifestándose si en P3 las calificaciones del grupo de autoevaluación se acercan más en correlación a las calificaciones provisionales del profesor en el grupo de autoevaluación en comparación con P1.

*RQ3*: No hay diferencias significativas entre las medias de calificaciones en ambos grupos otorgadas por parte del profesor (de forma provisional en el grupo de autoevaluación y definitivo en el de control).

Esta hipótesis nos permite analizar que el nivel de exigencia del profesor no se vea afectado por el experimento.

*RQ4*: No hay diferencia en los porcentajes de calificaciones y abandono de ambos grupos.

Esta hipótesis nos permite analizar cierto impacto de la autoevaluación en las calificaciones finales y la desmotivación (en forma de abandono) que se asocia tradicionalmente a las iniciativas de autoevaluación.

### 4. Resultados

Según las preguntas de investigación planteadas se realizaron los siguientes experimentos y consecuentes resultados.

RQ0: Se llevaron a cabo tests de dos muestras sobre igualdad de medias (Wilcoxon Rank Sum, H0: las medias son iguales, two-tail,  $\alpha=0.05$ ) [30, 31] en calificaciones de las parejas en ambos grupos, desagregando para cada práctica, en dos partes:

A- Wilcoxon Rank Sum en calificaciones finales del grupo de autoevaluación versus calificaciones finales del grupo de control por práctica son:

- P1 p-valor=0.36 (H0 no rechazada).
- P2 p-valor=0.01 (H0 rechazada)
- P3 p-valor=0.284284 (H0 no rechazada)

B- Wilcoxon Rank Sum en auto-calificaciones del grupo de autoevaluación versus calificaciones finales del grupo de control por práctica son:

- P1 p-valor=0.73 (H0 no rechazada).
- P2 p-valor=0.65 (H0 no rechazada).
- P3 p-valor=0.45 (H0 no rechazada).

Cuando analizamos las calificaciones que se dan los estudiantes a sí mismos con las calificaciones del grupo de control, no existe diferencia significativa en ninguna práctica. En el caso de la comparativa entre calificaciones finales de ambos grupos, existe una diferencia significativa en el caso de la práctica P2, mientras que en P1 y P3 no hay diferencias significativas en las calificaciones finales entre ambos grupos. Creemos que esto puede deberse puntualmente a la propia práctica P2 (en contenido, nivel de exigencia o dificultad para autoevaluar, todos estos criterios habrá que probarlos en el futuro), y no al efecto general de la autoevaluación.

RQ1: Se llevaron a cabo test de dos muestras sobre igualdad de medias (Wilcoxon Rank Sum, two-tail,  $\alpha=0.05$ ) [23, 26] de los incrementos (o decrementos) de calificaciones finales de ambos grupos de la práctica P1 a la práctica P3. Según el resultado (p-valor=0.32), no hay diferencia significativa en los incrementos entre grupos. Con los datos actuales no podemos por lo tanto concluir que el proceso de autoevaluación ayudase a mejorar más. Más experimentos deben realizarse para poder abordar los diferentes factores que estarían influyendo en este resultado.

RQ2: Se realizó un análisis de correlación Pearson entre las calificaciones establecidas por los estudiantes en el grupo de autoevaluación y las provisionales del profesor en ese mismo grupo, desagregado por práctica. La Figura 3 muestra las correlaciones obtenidas entre ambas calificaciones. Nótese que P1 (calificación final de P1), P1A (Autoevaluación de los estudiantes en P1) y P1P (Calificación profesor ciega

para P1) es la nomenclatura utilizada en todas las prácticas.

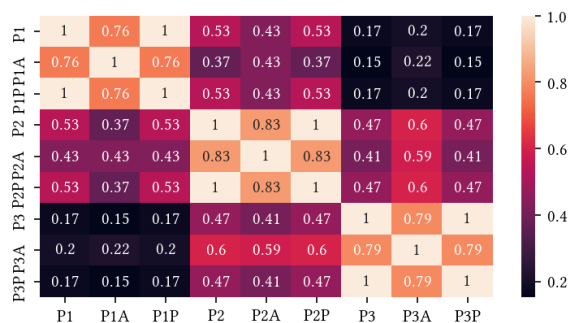


Figura 3: Correlaciones calificaciones grupo de Autoevaluación vs. calificaciones provisionales del profesor para ese mismo grupo.

Como puede observarse, la correlación entre la calificación del profesor y la autoevaluación cuando se evalúa la misma práctica (grupos centrales con resultados en naranja) va mejorando a lo largo del tiempo (hay mayor correlación en P2 (0.83) y P3(0.79) que en P1(0.76)). Esto nos da una noción preliminar acerca de que el grupo de autoevaluación muestra una mejora en cuanto a conciencia del propio proceso de aprendizaje después del primer ciclo, ya que en P2 y P3 las calificaciones del grupo de autoevaluación se acercan más en correlación a las calificaciones provisionales del profesor en ese grupo en comparación con P1. Un análisis de significancia específico con mayor muestra es necesario para confirmar este resultado positivo del protocolo diseñado.

RQ3: Se llevó a cabo un test de dos muestras sobre diferencia de medias (Wilcoxon Rank Sum, two-tail,  $\alpha=0.05$ ) [23, 26] en calificaciones finales provisionales del profesor en el grupo de autoevaluación y calificación final del grupo de control. Es decir, las calificaciones que otorga el profesor unilateralmente en ambos grupos. Según el resultado (p-valor=0.056), no variaría significativamente el nivel de exigencia del profesor entre grupos, si bien el p-valor se encuentra al borde la significancia. Analizando en profundidad las medias obtenidas, la conclusión inicial es que, dado que la labor del profesor es ciega a la calificación de autoevaluación, probablemente la diferencia se deba más a la naturaleza de cada grupo que a diferencias en el nivel de exigencia del profesor (ver RQ0). En cualquier caso, serían necesarios experimentos más específicos sobre el nivel de exigencia del profesor, aumentando el número de estudiantes y profesores, para analizar en profundidad el resultado de esta hipótesis.

RQ4: Se extrajeron los porcentajes por nivel de calificación y los porcentajes de abandonos (estudiantes que presentaron la P1 pero no la P2 o la P3) en

ambos grupos (autoevaluación y control), mostrados en Figura 4 y Figura 5 respectivamente.

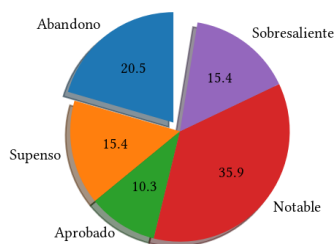


Figura 4: Porcentajes de calificaciones y abandono en el grupo de evaluación tradicional (Grupo de control).

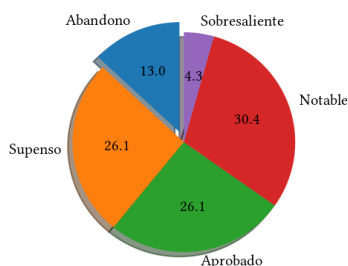


Figura 5: Porcentajes de calificaciones y abandono en el grupo de autoevaluación

Nótese en ambas imágenes dos aspectos de RQ4: 1) que el porcentaje de abandono es notablemente menor en el caso del grupo de Autoevaluación y 2) que, si bien se mantienen porcentajes similares, las calificaciones son más bajas en el grupo de Autoevaluación. En el primer caso, creemos que el protocolo de autoevaluación contribuye notablemente a reducir el porcentaje de abandono de los estudiantes, ya que ofrece retroalimentación detallada, aumenta la conciencia del propio proceso de aprendizaje (ver RQ2) y disuade del abandono a alumnos que tradicionalmente rechazan el sistema de defensa oral del método tradicional. En el segundo caso, creemos que el espectro de calificaciones en ambos grupos no difiere demasiado en representatividad. Esto, unido a la ausencia de rechazo de RQ0 nos lleva a pensar que las menores calificaciones en el grupo de autoevaluación puedan estar motivadas por otros factores no estudiados. Por ejemplo, la práctica docente nos hace pensar que el horario de las clases de los grupos de autoevaluación (principalmente por la tarde) puede ser un factor influyente en la bajada de las calificaciones generales del grupo, y no el protocolo de autoevaluación en sí.

## 5. Conclusiones

En resumen, los datos del primer curso académico de implantación del protocolo presentan resultados iniciales que creemos satisfactorios en RQ2 y RQ4,

que atienden a la mejora en la mejora incremental de la capacidad de autoevaluación y toma de conciencia del propio proceso de aprendizaje y a la reducción del porcentaje de abandono respectivamente. Estos dos factores son los que se ven claramente beneficiados con la aplicación del protocolo de autoevaluación y co-evaluación diseñado. Estos resultados están en la línea de los estudios previos que relacionan autoevaluación y coevaluación con menores tasas de abandono, más implicación por parte del alumnado y mejora de capacidades transversales. También ofrece resultados adecuados en cuanto a la no influencia del protocolo en el nivel de exigencia del profesor (RQ3), aunque son necesarios más experimentos para determinarlo de manera tajante. Sin embargo, este estudio empírico no ha encontrado diferencias significativas entre las calificaciones de autoevaluación y las del grupo de control, ni en las calificaciones finales (RQ0 y RQ1). Este aspecto puede presentar varias lecturas que deberán probarse con más experimentos. Por un lado, puede ser un resultado positivo si queremos implementar el protocolo como mecanismo alternativo de evaluación en la asignatura. Por otro, puede resultar negativo el hecho de que no consigamos mejorar las calificaciones de los estudiantes con el protocolo diseñado (RQ0), ni consigamos que las calificaciones mejoren incluyendo *feedback* propio (RQ1) del estudiante (asociado a la capacidad transversal de análisis crítico y detección de mejoras) en lugar del obtenido en defensa por parte del profesor (grupo de control).

Con respecto a la inclusión en el protocolo del aspecto de coevaluación en base a las parejas de prácticas, es importante destacar dos limitaciones: 1) ni el protocolo presentado ni el experimento diseñado para la evaluación de su implantación este curso 20/21 abordan los resultados de la autoevaluación de forma desagregada en caso de haber realizado cada miembro de la pareja la autoevaluación por separado y 2) estudios previos indicaban que, pese a sus buenos resultados en capacidades transversales y la buena aceptación por parte de los alumnos, los mecanismos de coevaluación arrojaban calificaciones de los alumnos “demasiado alejadas de las del profesor como para utilizarlas de manera rutinaria en la calificación de trabajos de respuesta abierta” [15].

En el primer caso, es posible hacerlo en el futuro, aunque se ha demostrado previamente que la coevaluación entre pares consensuada aumenta la precisión en la autoevaluación [19] ya que, una “corrección por parejas o en tríos resulta ligeramente más efectiva que una corrección individual, independientemente de la complejidad de la tarea” [19]. Estos estudios, unido a la estructura ya establecida de parejas en las prácticas de nuestra signatura de Sistemas Operativos, nos llevó a mantener esta estrategia en el protocolo de autoevaluación.

En el segundo caso, como puede verse en nuestros resultados, en nuestra experiencia no ha sido así. Creemos que el uso de una *check box* con criterios de evaluación claros por cada práctica evita las respuestas totalmente abiertas en la autoevaluación y co-evaluación y mejora la correlación entre las calificaciones del profesor y las de los alumnos.

## 6. Implicaciones futuras

El protocolo y primeros resultados de su aplicación constituyen tan sólo el primer paso de esta iniciativa docente en la asignatura de Sistemas Operativos. Como otros autores señalaron previamente [4], creemos necesario un análisis diacrónico de la implantación de dicho protocolo para poder extraer conclusiones acerca de la experiencia y primeros resultados aquí detallados, evitando el análisis de innovaciones docente cortoplacista [4]. Debido a esto, nuestro plan inmediato es continuar con la implantación del protocolo en sucesivos cursos académicos, balanceando más el número de alumnos en ambos grupos (el grupo de evaluación tradicional o de control y el grupo de autoevaluación), y variando aspectos que puedan influir y que se han detectado en algunas hipótesis, como el horario de clase del grupo o el profesor de prácticas involucrado en las sesiones de *Focus Group*. Creemos que esto puede suponer un ligero incremento de la dedicación docente en los próximos años por parte del profesorado. Sin embargo, una vez implantado, creemos que esta dedicación puede asimilarse a la dedicación empleada en la evaluación tradicional. La dedicación es también un aspecto a medir en futuras implantaciones.

Otro aspecto que reseñar es la especial relevancia que cobra este tipo de experiencias docentes en el contexto actual de entornos virtuales de aprendizaje [12]. A largo plazo, será necesario diseñar experimentos más específicos para medir el grado de mejora de cada capacidad transversal (análisis crítico, detección de posibles mejoras y conciencia del propio proceso de aprendizaje) por separado, con instrumentos específicos para cada una de ellas. En el contexto concreto de Sistemas Operativos, estos aspectos no han sido abordados de forma sistemática y cuantitativa en la literatura, siendo un área de trabajo futuro también para nuestro grupo de innovación docente.

## Referencias

- [1] Antonio J. Ballesteros Tallón. Enseñanza de ingeniería informática en inglés: sistemas operativos como caso de estudio. *Opción*, 31(1): p. 1175-1189, 2015.
- [2] Vigdis By Kampenes, Tore Dybå, Jo E. Hannay y Dag I.K. Sjøberg. A systematic review of effect size in software engineering

experiments. *Information and Software Technology*, 49(11-12): p. 1073-1086, 2007.

- [3] Ana Cadenato y María Martínez. L'avaluació en el marc de l'Espai Europeu d'Educació Superior (EEES). Monogràfics ICE de la Universitat Politècnica de Catalunya. Disponible en: [http://www.upc.edu/ice/portal-de-recursos/publicacions\\_ice/2\\_avaluacio.pdf](http://www.upc.edu/ice/portal-de-recursos/publicacions_ice/2_avaluacio.pdf), 2008.
- [4] Xavi Canaleta. Sistemas Operativos, un análisis a largo plazo de los procesos de innovación docente. En *Actas de las Jornadas sobre Enseñanza Universitaria de la Informática (JENUI) 2018*, 295-302, Barcelona, 2018.
- [5] María Cavas Toledo, Francisco Chicano García, Francisco Luna Valero y Luis Molina Tanco. Autoevaluación y evaluación entre iguales en una asignatura de redes de ordenadores. En *Actas de las XVIII Jornadas de Enseñanza Universitaria de la Informática (JENUI) 2011*, 303-310, Sevilla, 2011.
- [6] Agustín Cernuda del Río y Miguel Riesco Albizu. Una recapitulación sobre la autoevaluación de los alumnos en los estudios de informática: formas, utilidad y aplicación. En *Actas de las XVIII Jornadas de Enseñanza Universitaria de la Informática (JENUI) 2011*, 81-90, Sevilla, 2011.
- [7] María Ángeles Díaz Fondón, Miguel Riesco Albizu y Ana Belén Martínez Prieto. Hacia el aprendizaje activo: un caso práctico en la docencia de Sistemas Operativos. *Novática*, 174: p. 54-58, 2005.
- [8] Antonia Estero Botaro, Juan José Domínguez Jiménez y Manuel Palomo Duarte. Adaptación de las asignaturas Sistemas Operativos I y Sistemas Operativos II AL EEES. En *Jornadas sobre experiencias piloto de implantación del Crédito Europeo en las Universidades Andaluzas 2006*, 121-128, Cádiz, 2006.
- [9] Mercedes Fernández Redondo, Carlos Hernández Espinosa y Joaquín Torres Sospedra. Creación de herramientas de aprendizaje y autoevaluación para la asignatura sistemas operativos (IS11) de la ingeniería técnica en informática de sistemas. En *Metodologies centrades en l'estudiantat a l'Espai Europeu d'Educació Superior: actes de la VIII Jornada de Millora Educativa i VII Jornada d'Harmonització Europea de la Universitat Jaume I 2009*, 166-170, 2009.
- [10] Javier García Martín, Isabel Muñoz Fernández, Jorge Enrique Pérez Martínez y Trinidad Riolo Novalvos. Evaluación continuada en la asignatura de Sistemas Operativos I. En *Jornadas Nacionales de Intercambio de*

- Experiencias Piloto de Implantación de Metodologías ECTS. Aplicaciones Prácticas a la Convergencia Europea 2006, 1-9, Badajoz, 2006.
- [11] José Antonio Gómez y Buenaventura Clares, Experiencia en la adaptación de la asignatura Sistemas Operativos I al EEES, En Actas de las XIII Jornadas de Enseñanza Universitaria de la Informática (JENUI). 2007: Teruel. p. 73-80.
- [12] Jesús Enrique Londoño Salazar y Dora Aidé Ramírez González. Cultura de la autoevaluación: cimiento para la acreditación de programas académicos en modalidad virtual. Revista Virtual Universidad Católica del Norte,(36): p. 1-6, 2012.
- [13] Paula Lopez-Otero, Daniel Valcarce, Javier Parapar y Alvaro Barreiro. Docencia en Sistemas de Acceso a Información: detección de plaxios, empleo de tecnologías avanzadas para desenvolvimiento software e achegamento da experiencia na industria á aula. En Contextos universitarios transformadores: construíndo espazos de aprendizaxe. III Xornadas de Innovación Docente 2019, 41-56, A Coruña, 2019.
- [14] Eduardo Mosqueira Rey. La evaluación continua y la autoevaluación en el marco de la enseñanza de la programación orientada a objetos. En Actas de las XVI Jornadas de Enseñanza Universitaria de la Informática (JENUI) 2010, 223-230, Santiago de Compostela, 2010.
- [15] Javier Oliver Bernal y Verónica Canivell Castillo. Evaluación entre compañeros: estudio de su correlación con la evaluación del profesor. En Actas de las Jornadas de Enseñanza Universitaria de la Informática (JENUI) 2009, 241-246, Barcelona, 2009.
- [16] Sonia Pamplona Roche, Evaluación para el aprendizaje en la asignatura sistemas operativos en modalidad online: un estudio cualitativo basado en la taxonomía de Bloom. 2014, Tesis doctoral. ETSI\_Informatica. Universidad Politécnica de Madrid Madrid, España.
- [17] Jorge Enrique Pérez Martínez, Javier García Martín, María Isabel Muñoz Fernández, María Teresa Coello y Diana Pérez Martínez. Evaluación de competencias genéricas instrumentales en la asignatura de Sistemas Operativos I. En the 6th Simposio Iberoamericano en Educacion, Cibernetica e Informatica (SIECI) 2009, 1-4, Orlando (Florida), 2009.
- [18] Jorge Pomares Baeza, Pablo Gil Vázquez, Francisco A. Candelas-Herías, Santiago T. Puente Méndez, Fernando Torres Medina, Juan Antonio Corrales Ramón, Carlos A. Jara Bravo, Gabriel J. García Gómez y Iván Perea Fuentes. Sistema interactivo de aprendizaje autónomo y autoevaluación dentro del marco del EEES aplicado a la titulación ingeniería informática de la Universidad de Alicante En IX Jornades de xarxes d'investigació en docència universitària: disseny de bones pràctiques docents en el context actual 2011, 1605-1620, Alicante, 2011.
- [19] Juan Ramón Rico-Juan, Cristina Cachero y Hermenegilda Macià. Influencia de la evaluación entre pares consensuada en la precisión de las autoevaluaciones. En Actas de las XXVI Jornadas sobre la Enseñanza Universitaria de la Informática (JENUI) 2020, 205-212, 2020.
- [20] Miguel Riesco Albizu y M<sup>a</sup> Ángeles Díaz Fondón. Experiencia de uso de mapas conceptuales en la asignatura de Sistemas Operativos: dónde y cómo usarlos. En Actas de las XIV Jornadas de Enseñanza Universitaria de la Informática (JENUI) 2008, Granada, 2008.
- [21] Miguel Riesco Albizu y María Ángeles Díaz Fondón. Reinventando la rueda: una experiencia de aprendizaje por descubrimiento en la asignatura de Sistemas Operativos. ReVisión, 6 (1), 2013.
- [22] Dennis M. Ritchie. The development of the C language. ACM Sigplan Notices, 28(3): p. 201-208, 1993.
- [23] Sidney Siegel. Nonparametric statistics for the behavioral sciences. 1956.
- [24] Sílvia Terrasa Barrena y Gabriela Andreu García. Cambio a metodología de clase inversa en una asignatura obligatoria. En Actas del simposio-taller sobre estrategias y herramientas para el aprendizaje y la evaluación 2015, 32-37, Andorra La Vella, 2015.
- [25] Javier Traver y Juan Carlos Amengual. Una propuesta de autoevaluación-reflexión para potenciar la responsabilidad individual. En Actas de las Jornadas de Enseñanza Universitaria de la Informática (JENUI) 2013, 77-84, Castellón, 2013.
- [26] Frank Wilcoxon. Individual comparisons by ranking methods. Breakthroughs in statistics. 1992: Springer. 196-202.
- [27] Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, Björn Regnell y Anders Wesslén. Experimentation in software engineering. 2012: Springer Science & Business Media.