

La evaluación por pares en el ámbito universitario: una experiencia en el grado de ciencias de la actividad física y el deporte

Peer evaluation at the university setting: an experience in the degree of physical activity and sport

CARLES GONZÁLEZ-ARÉVALO

Instituto Nacional de Educación Física de Cataluña (INEFC)

Barcelona. España

cargonzalez@gencat.cat

ORCID: <https://orcid.org/0000-0003-1652-811X>

MERITXELL MONGUILLOT

Instituto Nacional de Educación Física de Cataluña (INEFC)

Barcelona. España

mmonguillot@gencat.cat

ORCID: <https://orcid.org/0000-0002-1114-6529>

ANA ANDRÉS

Facultad de Psicología, Ciencias de la Educación y del Deporte

Universidad Ramon Llull, Blanquerna, Barcelona. España

anaav2@blanquerna.url.edu

ORCID: <https://orcid.org/0000-0002-6488-1658>

Recibido: 25-10-2021. Aceptado: 22-11-2022.

Cómo citar / Citation: González-Arévalo, C., Monguillot, M. y Andrés, A. (2022). La evaluación por pares en el ámbito universitario: una experiencia en el grado de ciencias de la actividad física y el deporte, *Ágora para la Educación Física y el Deporte*, 24, 140-160.

DOI: <https://doi.org/10.24197/aefd.24.2022.140-160>

Resumen. El Espacio Europeo de Educación Superior supone un cambio de paradigma importante dando lugar a una enseñanza basada en el estudiantado donde la evaluación formativa cobra una especial relevancia. Esta nueva mirada plantea la evaluación como aprendizaje, promoviendo que el alumnado sea el protagonista de su propio proceso. El objetivo del estudio se centra en analizar las semejanzas y diferencias en la calificación emitida por el profesorado y por el alumnado en la modalidad doble ciega como estrategia para fomentar la autorregulación y mejora del aprendizaje del alumnado de segundo curso del Grado en Ciencias de la Actividad Física y el Deporte en el INEFC de Barcelona. El estudio se ha realizado a lo largo

de 5 cursos académicos en la asignatura de “Programación de la enseñanza de la actividad física y el deporte” y ha contado con una muestra de 453 alumnos y alumnas. El análisis longitudinal de los datos demuestra que no ha habido diferencias estadísticamente significativas entre los cinco cursos analizados aplicando el mismo sistema de evaluación. Los resultados obtenidos son satisfactorios y ponen de manifiesto el potencial de la evaluación por pares como estrategia de auto observación que facilita la autorregulación del aprendizaje del alumnado.

Palabras clave. Evaluación por pares; autorregulación; evaluación formativa; educación superior.

Abstract: The European Higher Education Area represents an important paradigm shift, giving rise to student-based teaching in which formative assessment takes on special relevance. This new perspective proposes assessment as learning, encouraging students to be the protagonists of their own process. The aim of the study is to analyse the similarities and differences in the marks awarded by teachers and students in the double-blind peer assessment as a strategy to promote self-regulation and improve learning among second-year students in the Bachelor's Degree in Physical Activity and Sport Sciences at the INEFC of Barcelona. The study was carried out over 5 academic years in the subject “Programming the Teaching of Physical Activity and Sport” and involved a sample of 453 students. The longitudinal analysis of the data shows that there were no statistically significant differences between the five academic years analyzed using the same assessment system. The results obtained are satisfactory and show the potential of peer assessment as a self-monitoring strategy that facilitates the self-regulation of student learning.

Keywords: Peer review; self-regulation; formative assessment; higher education.

INTRODUCCIÓN

La evaluación basada en el esquema tradicional de la enseñanza universitaria se ha entendido como una labor inherente al profesorado y se centra fundamentalmente en el examen escrito al final del período de docencia, donde lo que se mide principalmente son conocimientos, dejando muy poco margen para aprender de los errores. Sin embargo, las cuestiones para qué evaluar, qué evaluar, cuándo evaluar y quién evalúa admiten respuestas más amplias desde la óptica de la evaluación del aprendizaje (Brown y Glasner, 2003; Sanmartí, 2019). En esta línea, el marco del Espacio Europeo de Educación Superior (EEES en adelante), supone un cambio de paradigma importante dando lugar a una enseñanza basada en el estudiantado donde la evaluación formativa cobra una especial relevancia para la optimización del aprendizaje (Atienza et al., 2016; Calderón y Escalera, 2008; Fraile et al., 2018; Pérez et al., 2014; Quitiéiro, 2018). En consecuencia, en la actualidad se enfatiza un proceso de aprendizaje continuo y evolutivo como mecanismo de evaluación.

Más allá de la evaluación final, en este espacio se recomienda que los y las estudiantes reciban información sobre su progreso, sus deficiencias y si es posible orientaciones para la mejora a través del desarrollo de las distintas actividades propuestas. Esta información permitirá generar no sólo conocimiento sino mejorar y dotar de significado a todo el proceso de aprendizaje (Zubcoff et al., 2014). Así pues, el concepto de evaluación en la Educación Superior está cambiando, de un modelo centrado en el profesorado o de evaluación del aprendizaje hacia otro modelo centrado en el alumnado o de evaluación para el aprendizaje, haciendo hincapié en la evaluación centrada en las competencias de logro por parte de las y los alumnos que son los verdaderos protagonistas del proceso de enseñanza-aprendizaje (Pérez et al., 2014). Esta nueva mirada plantea la evaluación como aprendizaje, promoviendo que el alumnado sea el protagonista de su propio proceso siendo capaz de autoevaluarse e identificar qué sabe realizar, qué debe mejorar y cuál debe ser el siguiente paso (Sanmartí, 2019). En un contexto de reforma universitaria en el que los y las estudiantes han de desarrollar capacidades de pensamiento crítico mediante el estudio de las diferentes disciplinas es preciso introducir nuevos sistemas de evaluación que sean gratificantes, que mejoren la validez y fiabilidad de los métodos tradicionales e introduzcan en el proceso de evaluación la participación del propio estudiantado (Romero-Martín et al., 2015; Sanmartí, 2019; Valverde y Ciudad, 2014). En este contexto, conseguir que el alumnado aprenda es la gran finalidad de la enseñanza docente (Fraile et al., 2018; Sanmartí, 2019). Con este objetivo, será imprescindible delimitar las competencias; formular los objetivos de aprendizaje; definir los criterios de evaluación y de éxito, recoger las evidencias del desempeño que demuestran la adquisición de las competencias por parte del alumnado; analizar las evidencias para tomar decisiones sobre la mejora del aprendizaje logrado y comunicar los resultados a los y las alumnas para orientar y mejorar el proceso de aprendizaje (Sanmartí, 2019). Para ello, la evaluación deberá estar presente a lo largo de todo el proceso a través de diferentes estrategias e instrumentos de evaluación siendo llevada a cabo por el profesorado (heteroevaluación), el propio alumnado (autoevaluación) y los y las compañeras (evaluación entre iguales) (Gil y Padilla, 2009; Sanmartí, 2007). Por tanto, dentro de esta nueva concepción de la evaluación, cobra especial importancia el papel activo del propio alumnado como agente clave del proceso de aprendizaje. Una evaluación con carácter formativo debe tener en cuenta no solamente el producto final asociado al

rendimiento y calificación, sino también una evaluación que enfatiza el proceso que ha seguido el alumnado para llegar a la consecución de un producto final, y de las decisiones de mejora que ha ido tomando y que le han permitido aprender y conseguir el resultado final (Vergara, 2018). En esta nueva mirada de la evaluación, cobra especial protagonismo el *feedback*, entendido como la diferencia entre donde está el alumno al inicio de un proyecto o asignatura y dónde se pretende que esté al finalizar, es decir, entre los conocimientos previos o actuales y los criterios de éxito (Hattie, 2017). Esto implica permitir al alumnado, a través de las diferentes retroalimentaciones de su trabajo, ser más consciente, más autónomo y responsable de su proceso de aprendizaje. Tradicionalmente, esta retroalimentación suele recaer en la figura docente, lo que limita la reflexión del alumnado sobre su aprendizaje (Bordas y Cabrera, 2011; González et al., 2013). Es necesario, por tanto, que la evaluación sea en sí misma un aprendizaje, imprescindible para la adquisición de determinadas competencias genéricas a lo largo de los grados (Bautista-Cerro y Murga, 2011) y siendo el docente quien oriente, acompañe y guíe la construcción autónoma del aprendizaje del alumnado (Andreu et al., 2009).

De acuerdo con Sanmartí (2019), en el contexto de una evaluación para el aprendizaje, es fundamental que el alumnado aprenda a autorregularse. Para ello será importante que este sea capaz de representarse los objetivos de aprendizaje, anticipe y planifique las actividades a realizar y se apropie de los criterios de evaluación.

La evaluación para el aprendizaje se fundamenta en diferentes elementos como son la participación de los y las estudiantes, el *feedback* que permita la mejora, así como la incorporación de tareas auténticas e intencionales para desarrollar aprendizajes profundos (Atienza et al., 2016). Estos elementos ayudan a asumir procesos de evaluación y aprendizaje participativos, colaborativos basados en una concepción abierta, flexible y compartida del conocimiento. Para ello, será necesario desarrollar situaciones de aprendizaje basadas en el diálogo donde el alumnado va construyendo su conocimiento fruto de su discusión y reflexión con otras personas (Atienza et al., 2016; González, 2013).

Siguiendo a Sanmartí y Mas (2016) solo puede corregir errores la persona que los ha cometido, mientras que los y las compañeras, profesorado o familiares tienen la función de facilitar la corrección. Estas autoras sostienen que la evaluación entre iguales es muy útil puesto que identificar errores en otros facilita reconocer los que comete uno mismo.

En esta línea, la evaluación entre compañeros puede definirse como un cuerpo educativo donde el alumnado valora el desempeño de sus colegas cuantitativa y cualitativamente, estimulando de forma clave el aprendizaje autorregulado del estudiante (Butler y Winne, 1995).

Por otro lado, la autoevaluación y la evaluación entre iguales, además de permitir la evaluación de las competencias profesionales, fomenta y potencia otras competencias generales tales como el desarrollo de una visión crítica y reflexiva, la toma de decisiones y la asunción de responsabilidades, sin olvidar los valores de justicia y compromiso que subyacen en los procesos de evaluación entre iguales, con lo que se facilita la integración del alumno en el proceso de enseñanza-aprendizaje (Acedo y Ruiz-Cabestre, 2011; Gil y Padilla, 2009).

Son numerosos los trabajos que destacan la importancia de la evaluación por pares en el marco del EEES (Bautista-Cerro y Murga, 2011; Bernabé y Blasco, 2013). Varias revisiones de trabajos parecen este sentido (Falchikov, 2001; Topping, 1998) concluyen que existe una alta correlación entre las puntuaciones del alumnado y del profesorado. Incluso se apunta, a la luz de los resultados, que la evaluación entre pares se sugiere como un método de evaluación alternativo que puede ser aplicado de forma más frecuente en educación superior (Sahin, 2008). Por ello, la evaluación que deseamos, al margen de otras muchas posibles, está centrada y parte del propio alumnado, dándole voz para que provoque un aprendizaje más autónomo y profundo sobre el mismo proceso y acto de evaluación y, también, sobre las evidencias y criterios que se ponen en funcionamiento. Al mismo tiempo, nuestro alumnado se compromete e interesa no sólo por el mecanismo externo de la evaluación sino por su comprensión interna que le permite un aprendizaje continuo y verdadero a lo largo de la vida (Pérez et al., 2014).

Entre las ventajas que se le atribuyen a la evaluación por pares (Sahin, 2008; Sanmartí y Mas, 2016; Segers y Dochy, 2010; Topping, 1998), se destaca el identificar los errores y aciertos de los compañeros y en uno mismo, buscar estrategias para mejorar la calidad de los trabajos, aumentar la motivación del alumnado durante el proceso de aprendizaje y responsabilizarle de su evaluación. Se ha demostrado que la evaluación por pares es un buen predictor del compromiso académico y representa una oportunidad para potenciar el pensamiento crítico y reflexivo (Teich et al., 2015) y el trabajo en equipo (Oncu, 2015).

Sin embargo, son otras las desventajas de esta estrategia de evaluación, el alumnado puede no tener la madurez necesaria para ello y

no tomarse de forma seria la evaluación, mostrar actitudes negativas ante una evaluación de estas características o considerar la evaluación como una carga adicional (Brown y Glasner, 2003; Van den Berg et al., 2007). Además, hay evidencia que utilizar la calificación mediante la evaluación por pares, el alumnado suele puntuar más alto a sus compañeros que el propio profesor (Oncu, 2015).

Por último, recogemos las recomendaciones que hacen diversos autores con objeto de mejorar el proceso de evaluación por pares. Dochy et al. (1999) enfatizan la importancia que el proceso de evaluación sea anónimo (que no se conozca quién te evalúa ni a quién evalúas) a fin de garantizar que las relaciones sociales no afecten a la evaluación. La mayor parte de las evaluaciones entre pares realizadas se han basado en valorar trabajos en papel o en otros formatos que no impiden el anonimato de la persona a la que se evalúa, lo que aumenta la posibilidad de sesgo, la valoración parcial de amigos o el pacto entre compañeros en las calificaciones. No obstante, hoy en día, con las posibilidades técnicas de las que se dispone, esta desventaja puede ser mitigada de manera significativa.

En el presente estudio se ha aplicado la evaluación por pares en la modalidad doble ciega, como una modalidad que proviene del ámbito de la publicación de manuscritos, en la que ni quien revisa ni la persona autora se conocen y, por tanto, ambas son anónimas (Ladrón de Guevara et al., 2008). Para llevar a la práctica esta evaluación, destacamos la necesidad de reservar un tiempo adecuado y suficiente para que el alumnado pueda realizar sus valoraciones y usar los mismos criterios homogéneos para todos los agentes que intervienen en la evaluación, de forma que se determinen de antemano, sean conocidos por todos y los estudiantes estén ya familiarizados con ellos. Finalmente, se considera que la evaluación en el ámbito universitario necesita adaptarse a los intereses del alumnado y conectar con la realidad de la profesión. Con este objetivo, el alumnado necesita ser más autónomo y tener más protagonismo, ser más competente ante una realidad cambiante e incierta y organizar su propio proceso de aprendizaje. Paralelamente, el profesorado no puede seguir siendo un mero transmisor de conocimientos, sino que debe convertirse en un guía de aprendizaje que enseñe a filtrar contenidos y acompañe al alumno durante el proceso (Maya et al., 2020). De esta manera y de acuerdo con López-Pastor et al. (2005), consideramos la evaluación compartida y dialogada entre alumnado y profesorado como una estrategia ética y democrática que

implica al alumnado haciéndole partícipe y responsable de la toma de decisiones para la mejora de su aprendizaje.

1. OBJETIVO

El objetivo del estudio ha sido analizar las semejanzas y diferencias en la calificación emitida por el profesorado y por el alumnado en la modalidad doble ciega.

2. METODOLOGÍA

2.1. Participantes

El estudio se llevó a cabo durante 5 cursos académicos, con 453 alumnos y alumnas de 2º curso del grado de Ciencias de la Actividad Física y el Deporte (CAFD) del INEFC, centro de Barcelona, que cursaron la asignatura de “Programación de la enseñanza de la actividad física y el deporte”. Esta asignatura tiene una duración de seis créditos que se desarrollan mediante tres sesiones semanales de 1 hora y 30 minutos cada una. De ellas, dos son teóricas y el grupo lo forman unos 85 estudiantes, y una es práctica con unos 40 alumnos. El estudio se ha replicado durante los cursos académicos que abarcan el periodo del año 2013 al 2018.

2.2. Procedimiento

La asignatura cumple rigurosamente la condición de ser evaluada de forma continuada a lo largo del semestre mediante un total de 12 actividades de evaluación que combinan diferentes formatos: de heteroevaluación, de evaluación entre iguales y de autoevaluación, cediendo progresivamente al alumnado, el protagonismo en su proceso de evaluación. De estas 12 actividades, centraremos la atención en la corrección ciega de una unidad didáctica (en adelante UD). Esta actividad de evaluación es el foco del artículo que presentamos y la que ha sido objeto de la investigación en los cinco años de implantación del grado de CAFD del INEFC-centro de Barcelona. A continuación, en la tabla I mostramos las competencias básicas, específicas y los objetivos de aprendizaje que se desarrollan en la asignatura.

Tabla I. Competencias básicas, específicas y objetivos de aprendizaje de la asignatura

Competencias básicas	Competencias específicas	Objetivos de aprendizaje
<p>Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.</p> <p>Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía.</p> <p>Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio.</p>	<p>Planificar y evaluar los procesos y programas de la actividad física y del deporte en la enseñanza, entrenamiento, salud, gestión y recreación.</p>	<ol style="list-style-type: none"> 1. Analizar las fases del proceso de programación de una unidad didáctica. 2. Aplicar el proceso de programación en los diferentes ámbitos de la actividad física y el deporte. 3. Valorar la importancia de programar y los riesgos de no programar una unidad didáctica. 4. Programar una unidad didáctica contextualizada en un ámbito de la actividad física y el deporte específico.

2.3. Instrumento

El alumnado se ha distribuido por parejas y cada pareja debe programar una UD a lo largo del semestre. Esta actividad vale un 40% de la nota y es un ejercicio que simula una de las habilidades profesionales de más importancia: la capacidad de anticiparse y prever antes de actuar como enseñante. Para la realización de la UD, al inicio de la asignatura los y las estudiantes recibieron una base de orientación con unas detalladas instrucciones sobre cómo abordar cada fase, así como los criterios para su calificación. La base de orientación es un instrumento formativo de gran ayuda para el alumnado ya que le permitió anticipar, planificar y autorregular la calidad de su acción (Sanmartí, 2007) e ir confeccionando la UD de forma progresiva a lo largo del semestre. La UD estaba ponderada sobre 20 puntos y cada uno de ellos se correspondía con una de las fases que se recogen en la tabla II.

Tabla II. Fases de la unidad didáctica y puntuación

Fases de la programación de una UD	Puntos
Idea global de la programación	1
Diagnóstico inicial	2.5
Objetivos didácticos	6
Contenidos	0.5
Actividades de enseñanza-aprendizaje	1.5
Temporalización	0.5
Actividades y criterios de evaluación	7
Estrategias didácticas	0.5
Conclusiones	0.25
Bibliografía	0.25

Las fases de la programación de una UD fueron desarrolladas por cada pareja a la vez que fueron familiarizándose con los criterios de evaluación y calificación que recuperaron, más adelante, en la evaluación ciega. La tabla III recoge el instrumento para calificar la UD que ha utilizado el profesorado al finalizar del proceso.

Tabla III. Instrumento de evaluación y calificación de la UD

Presentación (Aspectos formales)
Encuadernación, formato, portada, título, índice, introducción, según instrucciones.
Idea global de la UD (1)
Explicad de forma resumida y redactada...
A quien va dirigida (población, edad...) (0.1)
Finalidad global de los aprendizajes) (0.1)
Tipos de objetivos (0.25)
Duración de la UD y momento de la temporada (0.1)
Tipos de actividades (0.1)
Opción metodológica (0.1)
Evaluación de los objetivos (0.25)
(Continúa)

Diagnóstico inicial (2.5)

Análisis del entorno (situación interna y externa) (0.5)

Marco teórico (mínimo de 5 referencias de autores/libros/estudios/entrevistas...) (0.5)

Características aprendices:

Datos personales y sociales + aspectos técnicos, tácticos y físicos (0.25)

Dos citas bibliográficas (0.25)

Evaluación inicial (prueba de evaluación inicial + criterios) (1)

Objetivos (6)

Si hay incoherencia entre la finalidad de la UD y los objetivos de la UD queda automáticamente suspendida

Frase inicial: "El aprendiz al acabar la UD será capaz de..." (0.25)

Formulados en una lista única + numerados + equilibrio entre conceptos/procedimientos/actitudes (0.25)

Dividir 5.5 entre el número de objetivos de la unidad

Multiplicar el número resultante sólo por el número de objetivos correctamente redactados

Contenidos (0.5)

Redactados con sustantivo y suficientemente desarrollados (0.25)

En una lista única, sin numerar y sin clasificar (0.25)

Actividades de aprendizaje (1.5)

Redactadas con sustantivo + de + actividad (1)

Cada actividad incorrectamente redactada resta -0.25.

Poner un ejemplo de tarea, juego, ejercicio, por actividad (0.25)

Listado único + numeradas + señalar expresamente las actividades de evaluación inicial, formativa y sumativa (0.25)

Temporalización (0.5)

Sólo en una hoja en un cuadro horario + número de sesión + objetivos numerados + contenidos y actividades de enseñanza aprendizaje numeradas: orden cronológico (0.25)

Sesiones de evaluación destacadas + añadir instrumento de evaluación subjetiva (0.25)

Actividades y criterios de evaluación (7)

3 condiciones que se deben cumplir:

Que estén valorados dentro de una actividad de evaluación (formativa o sumativa). Si no aparece la actividad penaliza la mitad del objetivo.

Que aparezca la dimensión cualitativa (aspectos observables en forma de rúbrica, lista de control, escala de valoración...).

Que aparezca la dimensión cuantitativa (se cuantifican los criterios cualitativos). Si no aparece, penaliza la mitad del objetivo.

Dividir 7 entre el número de objetivos de la UD = coeficiente máximo que supone cada objetivo valorado + multiplicar solo por los objetivos correctamente formulados si están valorados dentro de una actividad de evaluación y si poseen los criterios cualitativos de éxito y los cuantitativos.

(Continúa)

Estrategias didácticas (0.5)

1 sesión ejemplo: coherencia entre las partes de la sesión. (0.25)

Opción metodológica + justificación (global, analítica o/y mixta), + estilos enseñanza + justificación (producción o /y reproducción de modelos) + aspectos de organización de las sesiones (material, espacios y formación de grupos) (0.25)

Conclusiones (0.25)

De forma redactada: ¿Qué os ha aportado realizar la UD?, ¿Qué habéis aprendido?, ¿Os será útil?, ¿La pondréis en práctica?

Dificultades, sugerencias de mejora, grado de implicación entre la pareja y ayuda.

Bibliografía (0.25)

Deben aparecer las referencias que aparecen en el documento de acuerdo con la normativa APA 7.

2.4. Implementación de la evaluación y calificación por pares en la modalidad doble ciega

El procedimiento de implementación de la evaluación y calificación por pares en la modalidad doble ciega tuvo las siguientes fases:

- 1ª fase. Diseño del instrumento. Cada pareja diseñó su instrumento de evaluación y calificación de acuerdo con las indicaciones de la base de orientación. Este instrumento serviría para analizar la UD de otra pareja en la fase de la evaluación ciega.
- 2ª fase. Aplicación del instrumento. Cada pareja aplicó el instrumento diseñado para analizar su propia UD u otra que ya estuviera publicada. Esta fase permitió a las parejas detectar errores en el instrumento e introducir mejoras de cara a la evaluación ciega. El hecho de que cada pareja probase el instrumento de evaluación y calificación diseñado les permitió autorregular, corregir y mejorar su propia UD. Este hecho pone en valor el poder de la correulación como estrategia para favorecer la autorregulación, especialmente cuando se han compartido y consensado los criterios de evaluación (Sanmartí, 2007).
- 3ª fase. Validación del instrumento. En esta fase el profesorado validó los instrumentos de cada pareja a fin de prepararlos para ser aplicados en la fase de evaluación ciega, y ofreció *feedback* a aquellas parejas que necesitaban mejorar su instrumento.
- 4ª fase. Actividad de evaluación ciega. Esta actividad se realizó en clase. Cada pareja aplicó su instrumento para evaluar y calificar de forma anónima la UD de otra pareja escogida al azar a fin de

evitar cualquier tipo de sesgo (Martínez, 2012). De esta manera, se aseguró el anonimato y la corrección más “libre” y menos condicionada. Esta actividad era calificable y tenía un peso de un 10% en la nota final de la asignatura. Tener mejor o peor nota sobre el 10% dependía del acercamiento a la corrección realizada por el profesorado. Es decir, tanto profesorado como alumnado (en parejas) corregían la misma UD, y la calificación del alumnado dependía del acercamiento a la nota de la corrección realizada por el docente. Conforme las puntuaciones distan, la pareja va consiguiendo progresivamente menos nota conforme se aleja de la corrección del profesorado.

- 5ª fase. Elaboración del informe final de seguimiento. El proceso de evaluación y calificación finalizó con la realización por parte del alumnado de un informe final para valorar la importancia de la evaluación ciega y el proceso realizado hasta el día de su aplicación como estrategia de aprendizaje.

2.5. Recogida de datos

La recogida de datos se efectuó a lo largo de los cinco cursos académicos durante los meses de febrero a mayo. El instrumento utilizado por el profesorado para la recogida de datos fue una lista de control o check list tanto para la programación de la UD como para valorar y calificar el informe final. Los datos que se recogieron fueron las calificaciones de cada apartado de la UD y la nota final de la UD.

El instrumento que utilizó el alumnado para valorar y calificar la UD fue una lista de control que diseñó siguiendo las indicaciones de la base de orientación inicial. Además, el alumnado elaboró un informe de seguimiento cualitativo en equipo donde reflejó sus avances y dificultades durante el proceso de programación de la UD. El alumnado subió el informe a la plataforma Moodle de la asignatura.

2.6. Análisis de datos

El análisis de datos se realizó mediante el programa estadístico SPSS v.21. En primer lugar, se llevó a cabo un análisis descriptivo de los datos recogidos. Se analizó la normalidad de las variables mediante la prueba Kolmogorov-Smirnov, aplicándose las pruebas no paramétricas en caso

de que fuera adecuado. Concretamente, se calculó la correlación de Spearman para poner a prueba a la asociación entre variables cuantitativas. La comparación de medias para muestras relacionadas se llevó a cabo mediante la prueba T de Wilcoxon. Finalmente, las posibles diferencias en las evaluaciones realizadas a lo largo de los cursos se analizaron mediante la prueba no paramétrica Kruskal-Wallis.

Los datos cualitativos de los informes de seguimiento de cada equipo se organizaron en cuatro categorías: aspectos positivos, aspectos para modificar, sugerencias para mejorar la asignatura y valoración de la satisfacción global de la asignatura de 0 a 10.

3. RESULTADOS

3.1. Análisis descriptivo

En la tabla IV se muestran los datos estadísticos descriptivos de la calificación realizada por el profesorado y los propios alumnos y alumnas (por pares) a lo largo de los cinco cursos en los que se ha implementado este método de evaluación (n= 453).

Tabla IV. Estadísticos descriptivos de las calificaciones obtenidas a lo largo del tiempo

Cursos académicos	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015
Alumnado	n = 61	n = 92	n = 118	n = 102	n = 80
Calificación profesor					
Media (DT)	6.11 (1.92)	6.07 (2.16)	5.57 (2.01)	6.60 (1.68)	6.06 (1.99)
Mediana (RIC)	6.20 (3.50)	6.38 (3.25)	5.55 (3.38)	7.00 (2.29)	6.15 (2.98)
Calificación por pares					
Media (DT)	7.32 (1.35)	7.40 (1.72)	6.81 (1.73)	7.31 (1.59)	6.66 (1.87)
Mediana (RIC)	7.5 (2.20)	7.87 (2.04)	6.95 (2.50)	7.55 (2.28)	6.65 (2.83)

3.2. Análisis de correlaciones entre las calificaciones del profesorado y el alumnado

Con el objetivo de analizar la posible correlación entre las puntuaciones obtenidas a partir de la calificación del profesorado y de la calificación por pares, se llevó un análisis de correlaciones entre ambas

puntuaciones obtenidas por los 453 participantes. Tal y como indica el análisis de correlación, existe una asociación estadísticamente significativa entre ambas puntuaciones ($r=.466$, $p < .0001$). Por tanto, la calificación emitida por parte del profesorado covaría con la calificación por pares. Concretamente, la correlación osciló entre $.329$ y $.631$ en los diferentes cursos, siendo estadísticamente significativa en todos ellos. Sin embargo, el análisis de diferencias de medias pone de manifiesto que las puntuaciones obtenidas en la evaluación por pares fueron superiores a las obtenidas mediante la evaluación del profesorado ($Z=-10.472$, $p < .0001$).

3.3. Análisis de las diferencias en la calificación

A partir de las puntuaciones de calificación obtenidas por el profesor y los propios alumnos (pares) se calculó la diferencia entre ambas. Mediante la prueba no paramétrica de Kruskal-Wallis se puso de manifiesto que no había diferencias estadísticamente significativas en la diferencia de puntuaciones a lo largo de los cursos académicos analizados ($X^2 = 6.791$, $p = .079$).

La diferencia entre las puntuaciones fue transformada a puntuaciones normalizadas (z), poniéndose de manifiesto que estas diferencias oscilaron entre 2.58 y -3.43 desviaciones típicas por encima y por debajo de la media. Por tanto, se evidencia que en algunos casos se produce una discrepancia notable entre las puntuaciones de los dos evaluadores.

Se excluyeron las observaciones en las que se produjo una discrepancia entre observaciones mayor a $\pm 1z$, siendo el total de casos seleccionados $n=314$. El análisis de las diferencias de medias puso de manifiesto nuevamente diferencias estadísticamente significativas entre las calificaciones del profesor y las calificaciones por pares ($Z=-12.446$, $p < .0001$), mientras que no hubo diferencias entre los cursos ($X^2 = 4.836$, $p = .184$), tal y como ya se había constatado en este mismo análisis realizado antes de la transformación de las puntuaciones.

3.4. Análisis de los informes de seguimiento cualitativos del alumnado

De los informes de seguimiento cualitativos se desprenden los siguientes datos:

En relación con los aspectos positivos, el alumnado ha destacado el uso de una metodología participativa y diferente para aprender; las guías

e instrucciones brindadas le han ayudado en la resolución de las tareas, la relación de las sesiones teóricas con las prácticas le han facilitado la comprensión de los conocimientos; el uso de nuevas herramientas tecnológicas y el trabajo en equipo sistematizado le ha permitido aprender a cooperar.

El aspecto para modificar, y que mayoritariamente ha señalado el alumnado, ha sido el elevado volumen de trabajo a realizar lo que ha ocasionado una gran implicación fuera del horario lectivo.

Las sugerencias para mejorar la asignatura aportadas por el alumnado se han centrado en reducir el número de tareas o disponer de más tiempo para realizarlas y supervisar de forma más personalizada las tareas de evaluación entre iguales.

En cuanto a la valoración de la satisfacción global de la asignatura de 0 a 10, la nota mediana obtenida en los informes de todos los equipos ha sido de 8,1. Los comentarios del alumnado, que acompañan la nota, destacan la forma dinámica y proactiva de aprender y la gran inversión de tiempo que ha supuesto su realización.

4. DISCUSIÓN Y CONCLUSIONES

Una vez presentados los datos longitudinales (n=453) a lo largo de 5 cursos académicos, el procedimiento de evaluación y calificación utilizado no ha mostrado diferencias entre los 5 cursos en los que se ha aplicado. No obstante, a pesar de que existe relación entre la calificación del profesorado y la del alumnado, las calificaciones emitidas por el alumnado han sido superiores a las emitidas por el profesorado, y esta discrepancia se ha mantenido estable a lo largo de los cinco cursos en que se ha implementado la experiencia. Ante este resultado se observa cómo el juicio evaluativo (Fraile et al., 2020) del alumnado sobre la calificación de una UD todavía dista de la exactitud y experiencia del profesorado. A pesar de que el alumnado haya manifestado en los informes de seguimiento que disponer de los criterios de evaluación le ha permitido trabajar con exigencia los diferentes apartados de la UD: *“Entendemos que si no hubiera estos criterios de evaluación muchos grupos no trabajarían con tanta exigencia”*, los resultados arrojan a la luz la necesidad de cocrear los criterios de evaluación con el alumnado a fin de que éste se los haga suyos y desarrolle la capacidad para emitir un juicio evaluativo más preciso sobre la calificación de la UD. Sin embargo, y centrándonos en la parte formativa del proceso, se observa

cómo las actividades de enseñanza y aprendizaje implementadas antes de realizar la evaluación y calificación ciega han facilitado al alumnado apropiarse y manipular los criterios de evaluación para mejorar la calidad de su propia UD. De este modo, la autoevaluación, alejada de la calificación, se convierte en una potente herramienta para autorregular el aprendizaje del alumnado (Sanmartí, 2007). En esta línea el estudio ha contemplado cuatro estrategias para favorecer la autorregulación del alumnado.

En primer lugar, se ha proporcionado al inicio de la asignatura la base de orientación con las indicaciones sobre cómo realizar la UD.

En segundo lugar, se ha implicado al alumnado en la construcción de un instrumento de evaluación y calificación de la UD de acuerdo con los criterios indicados en la base de orientación.

En tercer lugar, el alumnado ha aplicado su instrumento sobre su propia UD u otra publicada a fin de validar y/o mejorar el instrumento antes de la evaluación ciega. Este hecho ha permitido al alumnado obtener un *feedback* efectivo, útil y con tiempo suficiente para ser usado en la mejora de la UD (Sadler, 2008).

Finalmente, el profesorado ha validado el instrumento y ha proporcionado un *feedback* de mejora antes de la evaluación ciega.

Este proceso sistemático ha permitido la autorregulación del alumnado mediante un *feedback* centrado en la tarea que ha permitido mejorar su propio aprendizaje.

Por otro lado, el hecho de compartir los criterios de evaluación e implicar al alumnado en la construcción del instrumento de evaluación y calificación que aplicaría posteriormente en la evaluación ciega, ha permitido al alumnado integrar los criterios de evaluación y ayudarle a comprender qué se exige en una determinada tarea (Johnson y Johnson, 2014). Los resultados de esta estrategia han sido satisfactorios. El alumnado ha conseguido los objetivos de aprendizaje vinculados con el diseño de una UD siendo protagonista de forma activa de su propio aprendizaje (Maya et al., 2020). El alumnado ha participado y ha supervisado su propio proceso de evaluación mediante la autoevaluación como estrategia y autorregulando su propio aprendizaje (Sanmartí, 2007).

En los informes de seguimiento, el alumnado ha valorado la experiencia como una de las más enriquecedoras en su formación, destacando sentirse mucho más autónomo y responsable a la hora de aprender a aprender. Algunos de sus comentarios así lo han manifestado:

“Manera original de aprender y hacer participe al alumnado de su propio proceso de aprendizaje.

En el presente estudio, la autosupervisión en base a unos criterios compartidos y conocidos y el *feedback* ofrecido y recibido por el docente con tiempo para la mejora durante el proceso, se han convertido en estrategias que han facilitado la autorregulación del aprendizaje del alumnado (Hattie, 2017; Panadero y Alonso-Tápias, 2014; Sanmartí, 2007).

Como posibles líneas de continuación del estudio, sería interesante profundizar en la evaluación entre pares en su sentido formativo, es decir alejado de la calificación, y entendiéndola como herramienta para el aprendizaje (Fraile, 2020; Hattie, 2017; Molloy et al., 2015). De igual modo, centrarse en la importancia del *feedback* entre iguales, no sólo a nivel de tarea sino también a nivel de proceso y de regulación abriría las puertas a descubrir estrategias para concienciar al alumnado sobre su propio aprendizaje.

BIBLIOGRAFÍA

- Acedo, M., & Ruiz-Cabestre, F. (2011). Una experiencia sobre la evaluación autónoma o participativa: autoevaluación y evaluación por los compañeros. *Arbor, Ciencia, Pensamiento y Cultura*, 187, 183-188. doi: <https://doi.org/10.3989/arbor.2011.Extra-3n3142>
- Andreu, Ll., Sanz, M.Á. & Serrat, E. (2009). Una propuesta de renovación metodológica en el marco del Espacio Europeo de Enseñanza Superior: los pequeños grupos de investigación cooperativos. *REIFOP*, 12 (3), 111-126. doi: <https://doi.org/10.4995/redu.2014.6415>
- Atienza, R., Valencia-Peris, A., Martos-García, D., López-Pastor, V. M., & Devís-Devís, J. (2016). La percepción del alumnado universitario de educación física sobre la evaluación formativa: ventajas, dificultades y satisfacción. *Movimento (ESEFID/UFRGS)*, 22 (4) 1033-1048. Doi: <http://dx.doi.org/10.22456/1982-8918.59732>
- Bautista-Cerro, M.J., & Murga, M.^a (2011). *La evaluación por pares una técnica para el desarrollo de competencias cívicas (autonomía y responsabilidad) en contextos formativos no presenciales. Estudio de caso.* XII Congreso de Teoría de la Educación. Barcelona: Universidad de Barcelona.

- Bernabé, G. & Blasco, S. (2013). Evaluación por pares y autoevaluación en el aula universitaria: una visión desde el enfoque por competencias. En M.T. Tortosa, J.D. Álvarez, & N. Pellín (Coord.) *XI Jornadas de redes de investigación en docencia universitaria* (pp. 2057-2069). Alicante: Universidad de Alicante.
- Bordas, M. & Cabrera, F. (2001). Estrategias de evaluación de los aprendizajes centrados en el proceso. *Revista Española de Pedagogía*, 218, 25-48.
- Brown, S., & Glasner, A. (Coord.). (2003). *Evaluar en la Universidad: problemas y nuevos enfoques*. Madrid: Narcea.
- Butler, D., & Winnie, P. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research*, 65(3), 245-281. Doi: <https://doi.org/10.2307/1170684>
- Calderón, C., & Escalera, G. (2008). La evaluación de la docencia ante el reto de Espacio Europeo de Educación Superior (EEES). *Educación XXI*, 11, 237-256.
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and coassessment in higher education: A review. *Studies in Higher Education*, 24, 331–350. doi: <https://doi.org/10.1080/03075079912331379935>
- Falchikov, N. (2001). *Learning Together: Peer Tutoring in Higher Education*. London: Routledge Falmer.
- Fraile, J., Pardo, R., & Panadero, E. (2018). Autoevaluación y autocalificación en el Grado en Ciencias de la Actividad Física y del Deporte: Estudio censal de las guías docentes. *Profesorado, Revista de Currículum y Formación de Profesorado*, 22(3), 163-182. doi: <https://doi.org/10.30827/profesorado.v22i3.7997>
- Fraile, J., Gil-Izquierdo, M., Zamorano-Sande, D., & Sánchez-Iglesias, I. (2020). Autorregulación del aprendizaje y procesos de evaluación formativa en los trabajos en grupo. *RELIEVE*, 26(1), art. M5. <http://doi.org/10.7203/relieve.26.1.17402>
- Gil, J., & Padilla, M.T. (2009). La participación del alumnado universitario en la evaluación del aprendizaje. *Educación XXI*, 12(1),43-65. doi: <https://doi.org/10.5944/educxx1.1.12.287>

- González, C., Navarro, I., Fernández, F., Heliz, J., Gilar-Corbi, R., Sánchez, B. Botella, P. (2013). La co-evaluación Una herramienta de aprendizaje. En M.T. Tortosa, J.D. Álvarez, & N. Pellín (Coord.) *XI Jornadas de redes de investigación en docencia universitaria* (pp. 568-583). Alicante: Universidad de Alicante.
- Hattie, J. (2017). *Aprendizaje visible para profesores*. Madrid: Paraninfo.
- Johnson, D., & Johnson, R. (2014). *La evaluación en el aprendizaje cooperativo*. Madrid: Ediciones SM.
- Ladrón de Guevara, M., Hincapié, J., Jackman, J., Herrera, O., & Vinicio, C. (2008). Revisión por pares: ¿Qué es y para qué sirve? *Salud Uninorte*, 24(2), 258-272. From: <https://www.redalyc.org/articulo.oa?id=81722411>
- López-Pastor, V.M., González, M., & Barba, J. (2005). La participación del alumnado en la evaluación: La autoevaluación, la coevaluación y la evaluación compartida. *Tándem: Didáctica de la Educación Física*, 17, 21-37.
- Martínez, G. (2012). La Revisión por pares y la selección de artículos para publicación. *Revista Colombiana de Psicología*, 21(1), 27-35.
- Maya, C., Iglesias, J., & Giménez, X. (2020). *Clase invertida síncrona en asignaturas STEM*. *Revista de educación*, 391,15-41. doi: <http://dx.doi.org/10.4438/1988-592X-RE-2021-391-469>
- Molloy, E., Borrell-Carrión, F., & Epstein, R. (2015). El impacto de las emociones en el feedback. En David Boud & Elizabeth Molloy (Eds.), *El feedback en educación superior y profesional. Comprenderlo y hacerlo bien* (pp.69-91). Madrid: Narcea.
- Oncu, S. (2015). Online peer evaluation for assessing perceived academic engagement. *Higher education. Eurasia Journal of Mathematics Science and Technology Education*, 11(3), 535-549. doi: <https://doi.org/10.12973/eurasia.2015.1343a>
- Panadero, E., & Alonso-Tápiá, J. (2014). ¿Cómo autorregulan nuestros alumnos? Revisión del modelo cíclico de Zimmerman sobre autorregulación del aprendizaje. *Anales de psicología*, 30 (2), 450-462. doi: <https://doi.org/10.6018/analesps.30.2.167221>

- Pérez, R., Cebrián, D., & Rueda, A.B. (2014). Evaluación de pares y autoevaluación con erúbricas: caso de estudio en el grado de Educación Primaria. *Revista de Docencia Universitaria*, 12(4), 437- 456. doi: <https://doi.org/10.4995/redu.2014.5632>
- Quitéiro, A.L. (2018). Assessment for learning in Physical Education. Its emergent relationships with motor competence development and intrinsic motivation toward physically literacy. *Ágora para la Educación Física y el Deporte*, 20(2-3), 213-234. doi: <https://doi.org/10.24197/aefd.2-3.2018.213-234>
- Romero-Martín, R., Castejón-Oliva, F.J., & López-Pastor, V.M. (2015). Divergencias del alumnado y del profesorado universitario sobre las dificultades para aplicar la evaluación formativa. *RELIEVE*, 21 (1). doi: <https://doi.org/10.7203/relieve.21.1.5169>
- Sadler, D. (2008). Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education*, 35(5), 535-50. From: <https://doi.org/10.1080/02602930903541015>
- Sahin, S. (2008). An application of peer assessment in higher Education. *The Turkish Online Journal of Educational Technology*, 7(2), 5-10.
- Sanmartí, N. (2007). *Evaluar para aprender*. Barcelona: Graó
- Sanmartí, N., & Mas, M. (2016). Les rúbriques per a una avaluació plantejada com a aprenentatge. *Revista Perspectiva Escolar*, 390, 37-41.
- Sanmartí, N. (2019). *Avaluar i aprendre: un únic procés*. Barcelona: Octaedro.
- Segers, M., & Dochy, F. (2010). New Assessment Forms in Problem-based Learning: the value added of the students' perspective. *Studies in Higher Education*, 26, 327-343. <https://doi.org/10.1080/03075070120076291>
- Teich, S., Demko, C.A., & Lang, L. (2015). Students' perception of peer-assessment in The context of a treatment planning course. *European Journal of Dental Education*, 19, 8-15. <https://doi.org/10.1111/eje.12095>
- Topping, K. (1998). Peer Assessment Between Students in Colleges and Universities. *Review of Educational Research*, 68, 249-276. <https://doi.org/10.4995/redu.2014.6415>

- Valverde, J., & Ciudad, A. (2014). El uso de e-rúbricas para la evaluación de competencias en estudiantes universitarios. Estudio sobre fiabilidad del instrumento. *Revista de Docencia Universitaria*, 12(1), 49-79.
- Van Den Berg, I., Admiraal, W., & Pilot, A. (2007). Design principles and outcomes of peer assessment in higher education. *Studies in higher education*, 31(3), 341-356. <https://doi.org/10.1080/03075070600680836>
- Vergara, J.J. (2018). *Narrar el aprendizaje*. España: Ediciones SM.
- Zubcoff, J.J., Forcada, A., Guardiola, J.V., & Gomariz, F. (2014). La evaluación por pares aplicado al proceso de aprendizaje en un escenario de evaluación continua. *XII Jornadas de Redes de Investigación en Docencia Universitaria. El reconocimiento docente: innovar e investigar con criterios de calidad* (pp. 2402-2416). Alicante: Universidad de Alicante.