

Participación del alumnado en la elaboración del examen teórico: análisis de resultados

Cristina Pascual-Arias¹; Miriam Molina-Soria²; Víctor M. López-Pastor³; David Hortigüela-Alcalá⁴

Recibido: Enero 2022 / Evaluado: Mayo 2022 / Aceptado: Septiembre 2022

Resumen. Los exámenes en la Formación Inicial del Profesorado (FIP) pueden repercutir tanto en la vida académica del alumnado como en su posterior docencia. Por ello, la aplicación de sistemas de Evaluación Formativa y Compartida (EFyC) pueden ayudar a mejorar su aprendizaje de manera constante. Se presenta un estudio con tres preguntas de investigación principales: (a) ¿Qué influencia puede tener, en la FIP, la implicación del alumnado en la elaboración de preguntas del examen en su rendimiento académico?; (b) ¿las calificaciones obtenidas en el examen final pueden verse influidas por la formación de procedencia de los estudiantes?; (c) ¿qué nivel de fiabilidad tiene la corrección del examen entre iguales respecto a la corrección del profesorado? La metodología empleada se basa en un estudio de caso con 35 alumnos divididos en dos grupos. Para la obtención de datos se emplea el acta con las calificaciones del examen. Los resultados muestran un alto rendimiento académico en la mayoría de los alumnos, pero no encontramos diferencias significativas entre grupos atendiendo a su formación de procedencia. También muestran una elevada y significativa fiabilidad en la corrección del examen entre alumnado y profesorado. Las implicaciones parecen positivas y relevantes en la FIP; ahora bien, parece conveniente comprobar los buenos resultados de este procedimiento con grupos que realicen modelos de examen más tradicionales, con muestras más amplias o con otras asignaturas de FIP.

Palabras clave: construcción de exámenes; evaluación compartida; evaluación formativa; formación inicial del profesorado; rendimiento académico.

[en] Student participation in the preparation on the theoretical examination: analysis of results.

Abstract. Examinations in the pre-service teacher education can have an impact on students' academic life as well as on their subsequent teaching. Therefore, the application of Formative and Shared Assessment (F&SA) systems can help to improve their learning on an ongoing basis. A study is presented with three main researched questions: (a) what influence can student involvement in the design of exam questions have on academic performance in the pre-service teacher education?; (b) can the grades obtained in the final exam be influenced by the students' background?; (c) how reliable is the peer making of the exam compared to teacher marking? The methodology used is based on a case study with 35 students divided into two groups. To obtain the data, the report with the exam marks is used. The results show a high academic performance in most of the students, but we did not find significant differences between groups according to their education background. They also show a high and significant reliability in the correction of the exam between students and teachers. The implications seem positive and relevant in the pre-service teacher education; however, it seems advisable to test the good results of this procedure with groups performing more traditional examinations models, with larger samples or with other pre-service teacher education subjects.

Keywords: test construction; shared assessment; formative assessment; pre-service teacher education; academic performance.

Sumario. 1. Introducción. 2. Metodología. 3. Resultados. 4. Discusión. 5. Conclusiones. 6. Referencias Bibliográficas.

Cómo citar: Pascual-Arias, C. et al. (2023). Participación del alumnado en la elaboración del examen teórico: análisis de resultados. *Revista Complutense de Educación*, 34(2), 379-388.

¹ Facultad de Educación de Segovia, Universidad de Valladolid (España)
E-mail: cristina.pascual@uva.es

ORCID: <https://orcid.org/0000-0002-2781-5600>

² Facultad de Educación de Segovia (Universidad de Valladolid) (España)
E-mail: miriam.molina@uva.es

ORCID: <https://orcid.org/0000-0003-2974-5535>

³ Facultad de Educación de Segovia, Universidad de Valladolid (España)
E-mail: victor.lopez.pastor@uva.es

ORCID: <https://orcid.org/0000-0003-2681-9543>

⁴ Universidad de Burgos (España)

E-mail: dhortiguela@ubu.es

ORCID: <https://orcid.org/0000-0001-5951-758X>

1. Introducción

1.1. Evaluación formativa y compartida en la formación inicial del profesorado

Desde hace más de dos décadas la educación universitaria está realizando el cambio necesario que planteó el Espacio Europeo de Educación Superior (EEES). Este cambio consiste en pasar del «Aprendizaje Bancario» hacia un modelo de «Aprendizaje Dialógico» (Fraile & Cornejo, 2012; López-Pastor, 2009). López-Pastor et al. (2020) señalan que el rol de la evaluación en el modelo de Aprendizaje Bancario se reduce a una medición final de resultados, frecuentemente mediante exámenes finales que se traducen a una única calificación final. En contraposición, el modelo de Aprendizaje Dialógico suele ir ligado a procesos de Evaluación Formativa y Compartida (EFyC), claramente centrados en la mejora de los aprendizajes (Andrade & Heritage, 2017; Dann & O'neil, 2017; López-Pastor, 2017; Moss & Bookhart, 2019). También puede ir ligado a otros cambios, como la utilización de metodologías y procesos de evaluación activos durante el proceso de enseñanza y aprendizaje, dentro de un enfoque más centrado en el desarrollo de competencias personales y profesionales (Boud & Falchikov, 2007; Brown & Glasner, 2003; López-Pastor et al., 2012; Nicol & Macfarlane, 2006).

En palabras de Dochy et al. (2002) este cambio en la evaluación significa pasar de una «cultura de examen» a una «cultura de la evaluación». Son muchos autores los que señalan que en la Formación Inicial del Profesorado (FIP) este cambio se debe producir dando un rol participante al alumno en su propia evaluación, tanto por el propio aprendizaje en sí, como para que posteriormente pueda aplicarlo en su futura docencia (Barberá, 2003; Biggs, 2005; Fraile, 2004; López-Pastor et al., 2012; Martínez, 2016). La participación del alumnado en su proceso de evaluación puede realizarse mediante técnicas como la autoevaluación, la evaluación entre iguales y la evaluación compartida (Dochy et al., 1999; Hortigüela-Alcalá et al., 2019; López-Pastor, 2009; López-Pastor & Sicilia, 2017).

López-Pastor (2009) considera que la EFyC es la metodología de evaluación más adecuada para dar respuesta al proceso de convergencia que propone el EEES, ya que se trata de un sistema orientado hacia la mejora de: (a) el aprendizaje del alumnado; (b) la docencia del maestro; (c) los procesos de enseñanza-aprendizaje que se llevan a cabo. En este proceso es imprescindible promover la implicación y participación del alumnado. Por tanto, hay varias razones para desarrollar la EFyC en la FIP: (a) la EFyC no sólo es un aspecto metodológico sino un contenido específico de aprendizaje y una competencia docente a desarrollar (López-Pastor et al., 2012); (b) la EFyC además de ayudar a mejorar muchas de las competencias docentes que se adquieren en la FIP, desarrolla la formación en evaluación como una competencia básica en sí misma (Gallardo-Fuentes, et al., 2018; Hortigüela-Alcalá et al., 2016; López-Pastor, et al. 2020; Martos et al., 2014; Romero et al., 2016; Silva & López-Pastor, 2015); (c) los docentes pueden realizar una transferencia entre la metodología de evaluación que vivieron en la FIP y su posterior práctica docente (Fraile, 2004; López-Pastor et al., 2012; Martínez, 2016); (d) la motivación, el aprendizaje y la implicación del alumnado mejoran notablemente (Barba-Martín & López-Pastor, 2017; López-Pastor et al., 2017); (e) la FIP tiene un fuerte componente práctico en consonancia con el aprendizaje dialógico que busca el EEES (López-Pastor, 2009).

1.2. Experiencias sobre usos formativos o innovadores del examen

El examen ha sido utilizado tradicionalmente como una prueba escrita, verbal, teórica o práctica para comprobar conocimientos memorísticos y como sistema de evaluación sumativa, orientado a la calificación del alumnado (López-Pastor et al., 2012; Pérez-Pueyo et al., 2017). El docente comprueba lo que se supone que debe saber el estudiante aplicando unos criterios más o menos conocidos a las respuestas efectuadas. El estudiante actúa de manera pasiva; su actividad queda limitada a responder y esperar el resultado de sus respuestas, primando el ámbito memorístico sobre cualquier otro.

Consideramos que para que el examen sirva como una herramienta más del proceso de aprendizaje y del proceso de EFyC es necesario utilizarlo fomentando la participación activa de los estudiantes, implicándolos en la corrección y elaboración de los exámenes. Pueden encontrarse algunas experiencias en esta línea en López-Pastor et al. (2012) o Pérez-Pueyo et al. (2017). Estos últimos autores señalan que para dar un uso formativo al examen hay diferentes posibilidades, por ejemplo: (a) utilizarlo como procedimiento de evaluación inicial o diagnóstica; (b) realizar procesos colectivos de corrección y revisión del examen tras su realización; (c) realizar procesos de autoevaluación o co-calificación.

En la literatura especializada pueden encontrarse algunos trabajos sobre la utilización formativa de los exámenes en FIP. El estudio de Pérez-Pueyo et al. (2014) es la base de la que parte la presente experiencia y en él se llevó a cabo una serie de fases para la elaboración y preparación de un examen, con sus respuestas y su posterior co-calificación y auto-calificación por parte del alumnado. Para realizar esta experiencia, Pérez-Pueyo et al. (2014) dividieron el total del grupo clase en sub-grupos de entre 6 y 8 miembros. Estos alumnos realizaron por grupos una batería de preguntas, tantas como integrantes tenía el grupo multiplicado por 10, con sus correspondientes respuestas sobre la asignatura siguiendo una serie de criterios concretos: (a) contener la referencia bibliográfica de dónde ha sido elaborada; (b) las respuestas tipo test debían tener cuatro

opciones; (c) las respuestas cortas no debían superar las cinco líneas; (d) las cuestiones debían tener el formato adecuado. Este proceso se llevó a cabo bajo la tutela del profesorado, que, a posteriori, eligió para el examen 30 preguntas tipo test y unas 8 o 10 preguntas cortas con las respuestas más interesantes presentadas. El proceso de elaboración de las preguntas también formó parte de la calificación del examen que quedó regulada: (a) calificación de la elaboración de las preguntas, 10% de la nota final; (b) calificación del examen, 30% de la nota final.

Los resultados del estudio muestran varias ventajas: (a) el alumnado adquiere un mayor aprendizaje, gracias a un mayor dominio de la materia; (b) el trabajo realizado motiva hacia la preparación del examen; (c) la revisión del trabajo ayuda a desarrollar capacidades cognitivas; (d) el alumnado tiene un proceso de aprendizaje más fructífero al manejar todo el material entregado.

Otra experiencia es la de utilizar el examen como una más de las diferentes actividades de aprendizaje y evaluación que se realizan en una asignatura, dentro de la FIP, pero con un peso pequeño en la calificación final (en torno al 20%), aunque hay que aprobarlo para superar la asignatura y, posteriormente, realizar un proceso de evaluación entre iguales, en el que los alumnos corrigen el examen de otro compañero con una plantilla de coevaluación. Pueden encontrarse diferentes estudios sobre este tipo de experiencias, con pequeñas diferencias en la realización, en los que los efectos positivos con respecto a realizar un examen tradicional son: (a) mejora el rendimiento académico del alumnado; (b) mejora la participación e implicación del alumnado; (c) mejora la adquisición de competencias docentes (Fraile et al., 2013; López-Pastor et al., 2012; Molina et al., 2019).

Destacamos un estudio en la misma línea llevado a cabo por López-Pastor et al. (2012) se basa en desarrollar el examen como una parte más de las actividades de evaluación de una asignatura en tres centros diferentes de FIP. El objetivo en este caso es comparar la calificación otorgada en la evaluación entre iguales y la calificación otorgada por el maestro, así como obtener las valoraciones del alumnado sobre esta experiencia. Los resultados indican que el alumnado otorgó calificaciones más bajas que el maestro, y que valoran positivamente la experiencia tanto para aprender como para ponerla en práctica como docentes. Podemos encontrar experiencias similares en etapas no universitarias, como el estudio de Lozano et al. (2017) en 1º de Educación Secundaria, con un grupo de 26 alumnos. En esta experiencia los alumnos elaboraron en grupos de tres o cuatro integrantes un total de 15 preguntas y sus repuestas, de las cuales el docente extrajo las más relevantes para la elaboración del examen. En esta experiencia el proceso de realización de las preguntas también formó parte del proceso de calificación: (a) 20% de la calificación para el funcionamiento del grupo, a través de un diario de seguimiento; (b) 30% de la calificación para la calidad de las preguntas recogidas; (c) 50% de la calificación para la nota individual obtenida en el examen. Los resultados muestran que las calificaciones del examen mejoraron de manera significativa, a la vez que incrementó el conocimiento del alumnado, su motivación e implicación en esta tarea. También surgió algún problema debido a la falta de costumbre y la reticencia de los alumnos al cambio de metodología para preparar el examen, ya que no se logró que todos los alumnos obtuvieran una calificación positiva en el mismo.

Debido a los beneficios que se han extraído de otras experiencias, consideramos que es pedagógicamente interesante plantear el examen como una actividad orientada hacia el aprendizaje del alumno, ya que parece aumentar su rendimiento académico. Este planteamiento deberá realizarse mediante un planteamiento en el que el alumnado participe en el proceso de elaboración y corrección del examen.

Por todo ello, las Preguntas de Investigación (PI) de este estudio son las siguientes:

- PI1-¿Qué influencia puede tener, en la FIP, la implicación del alumnado en la elaboración de preguntas de examen en su rendimiento académico?
- PI2-¿Las calificaciones obtenidas en el examen final pueden verse influidas por la formación de procedencia de los estudiantes (Grado en Educación Infantil vs. Doble Grado en Educación Infantil y Educación Primaria)?
- PI3-¿Qué nivel de fiabilidad tiene la corrección del examen entre iguales (co-evaluación) respecto a la corrección del profesorado?

2. Metodología

El diseño de esta investigación es un estudio de caso (Stake, 2005; Simons, 2011), que pretende analizar esta experiencia en concreto y describir esta realidad para generar conocimientos a partir de la misma. Con ello no pretendemos concretar resultados generales, sino detallar un proceso que pueda extrapolarse teniendo en cuenta la realidad de cada caso.

2.1. Participantes

Los participantes del estudio de caso son un grupo de 35 alumnos y alumnas de 4º curso de Grado de Educación Infantil de una universidad pública de la comunidad autónoma de Castilla y León. La muestra está compuesta

por los 35 alumnos que han elegido la vía continua de evaluación formativa y continua de la asignatura «Expresión y Comunicación Corporal en Educación Infantil», que se imparte en el primer cuatrimestre de cuarto curso de grado.

De los 35 alumnos, 33 son mujeres y 2 son hombres, con una media de edad de $22,85 \pm 1,58$ años. Pertenecen a dos grupos previos diferentes en el curso anterior: (a) Grupo 1: 18 alumnos procedentes de 3º curso del Grado de Educación Infantil; (b) Grupo 2: 17 alumnos procedentes del 4º curso del Plan de Estudios Conjuntos (Grado en Educación Infantil y Educación Primaria).

Los alumnos que proceden del Plan de Estudios Conjunto han realizado un año más de formación, sobre todo con las asignaturas correspondientes a la etapa de Educación Primaria. Se encuentran en quinto y último curso de su Doble Grado, en el que cursan las asignaturas optativas correspondientes al plan formativo de Educación Infantil. Además, han entrado en los estudios con notas de corte más altas, por lo que tienen un mayor nivel académico como punto de partida, además de un curso más de formación inicial y más asignaturas cada curso. Por tanto, parece lógico pensar que su nivel académico y formativo sea mayor que las que proceden de tercero de infantil.

2.2. Sistema de evaluación empleado

En la asignatura se ofrecen tres vías de aprendizaje y evaluación, para que cada alumno elija la que prefiera o que mejor se ajuste a sus circunstancias. Las tres vías son: (a) vía continua; los requisitos principales son la asistencia continua y la entrega de todas las actividades de aprendizaje programadas en tiempo y forma; (b) vía mixta, para los alumnos que asistan a más del 50% de las clases; (c) vía final, para los alumnos que no pueden asistir a clase; tienen que realizar 3 pruebas finales para poder ser evaluados y calificados.

La vía continua implica realizar un examen teórico al final del cuatrimestre que tiene un peso del 20% de la calificación final. Este estudio se centra en esta actividad concreta. Hay tres fases concretas en la realización de esta actividad de aprendizaje: (1) cada grupo de alumnos elaboran 8 posibles preguntas de examen y una rúbrica de corrección, con la supervisión y corrección del profesorado (ver descripción de este proceso en Tabla 1); (2) realización del examen teórico en la última semana lectiva; (3) corrección y evaluación entre iguales de un examen (co-evaluación), justo a continuación, y utilizando como plantilla las rúbricas de corrección. Los exámenes son también corregidos por el profesorado, por lo que de cada examen se dispone de dos puntuaciones: la del alumno que ha corregido y la del profesor.

2.3. Experiencia de elaboración y preparación de un examen de manera colaborativa

Para desarrollar la participación del alumnado en la preparación colaborativa del examen se tomó como planteamiento inicial las fases que plantean Pérez-Pueyo et al. (2014) y se transfirieron y adaptaron a este contexto concreto, como refleja la Tabla 1.

Tabla 1. *Fases detalladas de la experiencia de elaborar y preparar un examen en la FIP (elaboración a partir de Pérez-Pueyo et al., 2014)*

Fases de la experiencia	Características propias y detalle de las fases de elaboración y preparación del examen
Acotar el contenido del examen	El alumnado debía hacer las preguntas sobre los siguientes contenidos: (a) dossier de teoría elaborado en clases de la asignatura. (b) textos sobre los que se habían hecho tertulias dialógicas en los seminarios. (c) marcos teóricos de los PAT elaborados por cada grupo.
Elaboración de preguntas	Cada uno de los 12 grupos tuvo de 3/5 miembros, y realizó 8 preguntas sobre el contenido de examen. Para cada pregunta se acordó realizar una rúbrica de corrección con 4 niveles, para realizar la posterior corrección del mismo. Los principales criterios para la elaboración de las preguntas fueron: (a) las preguntas deben incitar a la reflexión de todos los contenidos de la asignatura. (b) las preguntas deben contener al menos dos de las siguientes acciones: recordar, aplicar, comprender, analizar, sintetizar, realizar, valorar, programar o expresar. (c) las propuestas deben generar respuestas reflexivas, interesantes y adecuadas que permitan comprobar la adquisición de los contenidos de la asignatura. (d) todas las respuestas fueron acompañadas de una rúbrica de corrección con cuatro niveles de logro y sus ítems correspondientes.
Selección de preguntas	De las 96 preguntas totales los docentes de la asignatura seleccionaron 48 preguntas, tras eliminar las incorrectas, las incoherentes y las repetidas entre grupos. Estas 48 preguntas pasaron a formar parte del banco de preguntas que se le ofreció al alumnado para preparar el examen.

Elaboración de modelos de examen	Elaboramos dos modelos de exámenes (A y B) con 7 preguntas similares en cada uno, extraídas de las 48 preguntas totales. Junto a los modelos de examen se realizaron las plantillas de corrección con las rúbricas que habían realizado los grupos. Las preguntas del examen las seleccionaron los docentes de la asignatura para que el examen recogiera todo lo trabajado en la asignatura de la siguiente forma: (a) 3 preguntas de los PAT elaborados en la asignatura. (b) 2 preguntas de la parte teórica. (c) 2 preguntas de la parte de lecturas de seminario.
Desarrollo y corrección del examen	En esta fase encontramos grandes diferencias, los alumnos realizaron el examen durante 1 hora de tiempo. Después cada alumno corrigió el examen de otro compañero con la rúbrica de corrección que habían elaborado para cada pregunta. Facilitar la rúbrica de corrección afianza la objetividad, puesto que los exámenes tienen el nombre del alumno. La calidad de la corrección del examen de otro compañero tiene consecuencias en la nota del examen de cada alumno, con tres posibilidades: (a) Si el alumno corregía con una diferencia de máximo 0.10 puntos con respecto a la nota que ponía el profesor, obtenían 0.10 más en su examen. (b) si el alumno corregía con una diferencia de entre 0.10 y 0.20 puntos con respecto a la nota que ponía el profesor, no tenían repercusión en su nota de examen. (c) si el alumno corregía con una diferencia de más de 0,20 puntos con respecto a la nota que ponía el profesor, se le restaba 0,10 de su examen. El proceso de elaboración de preguntas no forma parte de la nota del examen.

2.4. Técnicas de obtención de datos

Para la obtención de datos se han utilizado las siguientes técnicas: (a) acta con las calificaciones del examen obtenidas por el alumnado que opta por la vía continua de evaluación; (b) hoja de registro con las notas del alumnado en la corrección del examen de un compañero.

2.5. Técnicas de análisis de datos

El proceso de análisis de datos se ha organizado para dar respuesta a cada una de las PI, aplicando para ello el programa estadístico SPSS 24.0. Para resolver la PI1 se ha realizado un análisis descriptivo, utilizando para ello la distribución de porcentajes en cada rango de calificación, la obtención de la media aritmética (M) y la desviación típica (DT.), en cada uno de los grupos de procedencia. Para la PI2 se ha aplicado la prueba de estadística inferencial T de Student, para comprobar si hay diferencias significativas entre la calificación media de los alumnos en función a su grupo de procedencia. Para la PI3 se ha aplicado la correlación de Pearson al ser una muestra paramétrica, para comprobar si hay una correlación estadísticamente significativa entre las calificaciones otorgadas por los alumnos y por los docentes.

En cuanto a los criterios éticos de la investigación se han tenido en cuenta una serie de cuestiones como el respeto por la autonomía de los participantes y, durante el tratamiento y análisis de datos, se ha asegurado el anonimato y la confidencialidad de los participantes (Bisquerra, 2004).

3. Resultados

Hemos dividido los resultados en tres grandes categorías, atendiendo a cada una de las tres preguntas de investigación propuestas: (a) influencia en el rendimiento académico de la participación del alumnado en la elaboración de las preguntas del examen; (b) diferencia de rendimiento entre los dos grupos de procedencia; (c) nivel de fiabilidad de las correcciones otorgadas por los alumnos.

3.1. Influencia en el rendimiento académico de la participación del alumnado en la elaboración de las preguntas del examen

La primera pregunta de investigación plantea qué influencia puede tener, en la FIP, la implicación del alumnado en la elaboración de preguntas de examen en el rendimiento académico. En la tabla 2 puede observarse como la mayoría del alumnado supera el examen y como la calificación media es alta (7.78 sobre 10). El porcentaje de suspensos es bajo (8.44) y los porcentajes más altos se dan entre notable y sobresaliente.

Tabla 2. *Calificación media obtenida en el examen y porcentajes por rango de calificación.*

Calificación obtenida	Porcentaje de alumnos
Sobresaliente	37.41
Notable	31.37
Aprobado	22.78
Suspenso	8.44
Calificación media	7.78
Desviación típica	1.84

3.2. Diferencia de rendimiento entre los dos grupos de procedencia

La segunda pregunta de investigación plantea si las calificaciones obtenidas en el examen final pueden verse influidas por la formación de procedencia de los estudiantes (3º de Infantil vs. 4º de doble grado). Se presentan los resultados en las tablas 3 y 4. Los datos de la tabla 3 muestran que el rendimiento es claramente mejor en el grupo 2: mayor porcentaje de sobresalientes, ningún suspenso y mayor media. Por tanto, los datos parecen indicar que el grupo 2 obtuvo mejores calificaciones en general que el grupo 1; ahora bien, hay que comprobar si dichas diferencias son estadísticamente significativas.

Tabla 3. *Calificaciones obtenidas en el examen, porcentajes por rango y calificación media diferenciando entre los grupos de procedencia del alumnado.*

Calificación obtenida	Grupo 1: Alumnos procedentes de 3º de Educación Infantil	Grupo 2: Alumnos procedentes de 4º de Doble Grado de Educación Infantil y Educación Primaria
	Porcentaje	Porcentaje
Sobresaliente	27.78	47.05
Notable	33.33	29.41
Aprobado	22.22	23.34
Suspenso	16.67	0
Media	7.39	8.19
Desviación típica	1.97	1.65

Dado que se observan diferencias en las medias de ambos grupos, en la tabla 4 presentamos el resultado de la prueba T de Student, que nos permite comprobar si la diferencia de medias es estadísticamente significativa.

Tabla 4. *Diferencia de medias entre los grupos de procedencia*

	Calificación Media	Desviación típica	Diferencias significativas prueba T de Student (Sig. Bilateral)*
Grupo 1: Alumnos procedentes de 3º de Grado de EI	7.39	1.97	.203
Grupo 2: Alumnos procedentes de 4º de Grado de EI y EP	8.19	1.65	

Nota: * Valor de significancia $p \leq .05$ (bilateral)

No se han encontrado diferencias significativas entre ambos grupos de procedencia.

3.3. Nivel de fiabilidad de las correcciones del examen realizadas por los alumnos

La tercera pregunta de investigación hace referencia a qué nivel de fiabilidad tiene la corrección del examen entre iguales (co-evaluación) respecto a la corrección del profesorado (heteroevaluación). En la tabla 5 mostramos los resultados de las correcciones de exámenes entre iguales (en escala de calificación 0-10), los resultados de la calificación del docente y la correlación entre ambas en cada uno de los dos grupos de procedencia.

Tabla 5. *Correlación entre los resultados de la co-calificación de los alumnos y la calificación de los docentes del examen preparado colaborativamente.*

	Calificaciones asignadas por los alumnos (co-evaluación)	Corrección de los profesores (heteroevaluación)
Resultados por rango de calificación	Sobresaliente: 13	Sobresaliente: 13
	Notable: 11	Notable: 11
	Aprobado: 11	Aprobado: 8
	Suspense: 0	Suspense: 3
Correlación de Pearson*	.872	
Sig. (bilateral)	<.001	

Nota: * La correlación es significativa en el nivel 0.01 (bilateral)

Los resultados muestran una correlación muy alta y estadísticamente significativa, entre la evaluación que realizan los alumnos y la que realiza el docente. En cuanto a los resultados mostrados de “calificación entre iguales y calificación de los docentes”, se observa que son prácticamente iguales, se han otorgado el mismo número de sobresalientes que de notables. Destaca como diferencia que el alumnado no ha evaluado ningún examen con la calificación de suspenso, mientras que, en la calificación de los docentes, encontramos 3 alumnos que no han superado el examen.

4. Discusión

Se han analizado los resultados de un examen implicando al alumnado en su elaboración y corrección en la FIP. Tras desarrollar un examen de manera colaborativa implicando al alumnado, en consonancia con el proceso de EFyC en el que se basa esta asignatura, la elaboración del examen ha sido mucho más sencilla, adaptando las fases propuestas por Pérez-Pueyo et al. (2014).

La primera pregunta planteaba qué influencia puede tener, en la FIP, la implicación del alumnado en la elaboración de preguntas de examen en el rendimiento académico. Los resultados muestran un elevado rendimiento académico en el examen, con un bajo número de suspensos y un elevado porcentaje de sobresalientes y notables. Ahora bien, no podemos asegurar que sea debido a la preparación de preguntas por parte del alumnado, porque el diseño de investigación no permite asegurar que sea la única variable que está influyendo. Estos resultados coinciden con los encontrados en otros estudios, que muestran que, en las ocasiones en las que se ha medido (en base a los resultados obtenidos) cuando el alumnado participa en el proceso de elaboración de preguntas para el examen parece que aprende más y, en consecuencia, obtiene mejores resultados (Lozano et al., 2017; Pérez-Pueyo et al., 2014), aunque en estudios comparados entre centros, pueden encontrarse diferencias considerables en función de la asignatura y centro (Pérez-Pueyo et al., 2014).

La segunda pregunta de investigación se centra en si las calificaciones obtenidas en el examen final pueden verse influidas por la formación de procedencia de los estudiantes (3º curso del grado de Educación Infantil vs. 4º curso del Doble Grado de Educación Infantil y Educación Primaria). Como se ha podido comprobar en el apartado de resultados, si que se observan algunas diferencias entre ambos grupos (en porcentaje de sobresalientes y suspensos y en la media), pero no son estadísticamente significativas. Otras diferencias que hemos encontrado son: (a) la nota media en el examen es mayor en el grupo 2 que en el grupo 1, con casi un punto de diferencia. Este hecho puede deberse a que han recibido un año más de formación o a otras experiencias similares, como hemos destacado en el párrafo anterior, pero no tenemos evidencias sobre este aspecto; (b) en el grupo 2 no ha habido suspensos en la primera convocatoria, mientras que en el grupo 1 hubo 3 que en la segunda convocatoria aprobaron con calificaciones de bien y notable. En este sentido, estudios como los de Gallardo et al. (2018), Hortigüela-Alcalá et al. (2015) o Buscá et al. (2012) muestran que los alumnos de cursos superiores tienen una mayor implicación en su aprendizaje, aspecto que podría explicar que en este estudio hayan obtenido mayores puntuaciones.

La tercera pregunta de investigación se centra en qué nivel de fiabilidad tiene la corrección del examen entre iguales (co-evaluación) respecto a la corrección del profesorado (heteroevaluación). Como se ha comprobado, hay una alta correlación entre la calificación realizada por los alumnos y por los docentes. Esto puede deberse al proceso de revisión y feedback que se realiza de manera conjunta entre los alumnos y los docentes, así como que los alumnos habían preparado y elaborado previamente las rúbricas de corrección, desarrollando así competencias profesionales propias de la FIP, como razonar de forma reflexiva y crítica, desarrollar un aprendizaje autónomo o afrontar deberes y dilemas éticos de la profesión, al tener que evaluar a sus compañeros con rigor (Lozano et al., 2017).

Estos resultados parecen estar en la línea de lo que señalan López-Pastor et al. (2012), al destacar que cuando los criterios de calificación son claros, las fluctuaciones en las notas entre las valoraciones del profesor

y del alumno son pequeñas. También se puede hacer alusión a estudios como los de Boud (1995, 2020), Boud & Fachinov (2006) y Falchikov (2005), que afirman que el curso, el nivel académico y la madurez personal son variables que afectan a la fiabilidad de los procesos de autoevaluación o evaluación entre iguales en alumnos universitarios.

5. Conclusiones

Los resultados parecen indicar que esta experiencia aporta beneficios relativos en el rendimiento académico de los alumnos, ya que la gran mayoría de los alumnos han obtenido calificaciones altas en el examen, aunque no podemos garantizar que sea la única variable que está influyendo.

Comparando por formación de procedencia, los alumnos del Doble Grado obtienen calificaciones más altas que los alumnos del Grado de Educación Infantil, aunque estas diferencias no son estadísticamente significativas.

Existe una correlación muy alta y estadísticamente significativa entre la corrección del examen realizada por los alumnos y la realizada por los docentes, lo que aporta fiabilidad al proceso y refuerza el desarrollo de las competencias profesionales relacionadas con la evaluación y calificación en la FIP.

En este proceso hemos revisado estudios similares y creemos que debemos considerar para futuras aplicaciones la preparación de las preguntas del examen como una actividad de aprendizaje y de calificación más. De esta manera, el proceso de preparación de las preguntas del examen también puede ser calificado dentro de la nota global del examen, dando más sentido a la realización del examen como actividad de aprendizaje.

Este trabajo puede resultar de interés para el profesorado de FIP por varios motivos: (a) puede ayudar a generar y afianzar en los alumnos de FIP competencias específicas docentes relacionadas con la evaluación; (b) esta experiencia puede extrapolarse a otras etapas educativas y, por tanto, los alumnos que la vivencien en la FIP pueden ponerla en práctica en su futura docencia; (c) puede servir para mejorar la participación e implicación del alumnado en su propio aprendizaje en la FIP. También puede ser de interés para docentes de otras etapas educativas que quiera realizar una práctica de examen de manera participativa y colaborativa con el alumnado.

Los hallazgos más importantes de este estudio podrían sintetizarse en que hacer partícipe al alumnado de la elaboración y corrección del propio examen es posible y tiene varias ventajas, especialmente en la FIP. Además, puede incidir positivamente en su aprendizaje y dominio de la materia, les permite desarrollar competencias docentes relacionadas con la evaluación y la calificación, que pueden serles muy útiles en su futuro profesional. Sobre las limitaciones del estudio, resalta el hecho de que se ha llevado a cabo con un sólo grupo-clase en un único contexto, por lo que se puede asegurar la influencia única o principal de la variable estudiada.

Por todo ello, la prospectiva de futuro muestra dos vías de acción. La primera de carácter didáctico, intentando mejorar la actividad a partir de las propuestas de mejora realizadas por el alumnado y profesorado de la asignatura. La segunda, de carácter investigador, iría dirigida a comparar resultados de experiencias de elaboración y corrección de examen con otros grupos que puedan realizar esta experiencia en otras asignaturas o contextos, así como ampliar esta experiencia en la misma asignatura y comprobar sus resultados.

6. Referencias Bibliográficas

- Andrade, H. L. & Heritage, M. (2017). *Using Formative Assessment to Enhance Learning, Achievement and Academic Self-Regulation*. Routledge
- Barba-Martín, R. & López-Pastor, V. (2017). Evaluación formativa y compartida en los procesos de trabajo tutorado, un ejemplo de buena práctica. *Revista Infancia, Educación y Aprendizaje*, 3(2), 66-70. <https://doi.org/10.22370/ieya.2017.3.2.701>
- Barberá, E. (2003). Estado y tendencias de la evaluación en educación superior. *Revista de la Red Estatal de Docencia Universitaria*, 2(3), 94-99. <https://revistas.um.es/redu/article/view/10161>
- Biggs, J. (2005). *Calidad del aprendizaje universitario*. Narcea.
- Bisquerra, R. (2004). *Metodología de la investigación educativa*. La Muralla.
- Boud, D., (ed.). (1995). *Enhancing learning through self- assessment*. Routledge Falmer.
- Boud, D. (2020). Retos en la reforma de la evaluación en educación superior: una mirada desde la lejanía. *RELIEVE*, 26(1). <https://doi.org/10.7203/relieve.26.1.17088>
- Boud, D. & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment and Evaluation in Higher Education*, 3(4), 399-413. <https://doi.org/10.1080/02602930600679050>
- Boud, D. & Falchikov, N. (2007). *Rethinking Assessment in Higher Education. Learning for the long term*. Routledge.
- Brown, S. & Glasner, A. (2003). *Evaluar en la Universidad. Problemas y nuevos enfoques*. Narcea.
- Buscá, F., Rivera E. & Trigueros, C. (2012). La credibilitat dels sistemes d'avaluació formativa en la formació inicial del professorat d'educació física. *Temps d'Educació*, 43, 167-184. <http://diposit.ub.edu/dspace/handle/2445/113262>

- Castejón, F. J., López-Pastor V. M., Julián, J. A., & Zaragoza, J. (2011). Evaluación Formativa y rendimiento académico en la formación inicial del profesorado de Educación Física. *Revista Internacional de Medicina y Ciencias de la Actividad Física y del Deporte*, 11(42), 238-346. <https://bit.ly/2IUMkml>
- Dann, C. & O'Neill, S. (2017). Positioning Preservice Teacher Formative Assessment in the Literature. En T. Richardson, D. Beverly, C. Dann & S. O'Neill (Eds.), *Formative Assessment. Practices for pre-servide Teacher Practicum Feedback: Emerging Research and Opportunities* (pp.1-46). IGI Global.
- Dochy, F., Segers, M. & Dierick, S. (2002). Nuevas vías de aprendizaje y enseñanza y sus consecuencias:, una era de evaluación. *Red Estatal de Docencia Universitaria*, 2(2), 13-30. <https://bit.ly/2IU0Xqs>
- Dochy, F., Segers, M. & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: a review. *Studies in Higher Education*, 24, 331-350. <https://doi.org/10.1080/0375079912331379935>
- Falchikov, N. (2005): *Improving Assessment Through Student Involvement. Practical solutions for aiding learning in higher and further education*. Routledge-Falmer.
- Fraile, A. (2004). Un cambio democrático en las aulas universitarias: una experiencia en la formación del profesorado de Educación Física. *Contextos educativos: Revista de Educación*, 6(7), 213-234. <https://bit.ly/2U92y0S>
- Fraile, A. & Cornejo, P. (2012). La evaluación formativa en la enseñanza universitaria: una experiencia de innovación educativa con estudiantes de Educación Física. *Revista de evaluación educativa*, 1(2). <https://bit.ly/2IKNJwn>
- Fraile, A., López-Pastor, V.M., Castejón, J. & Romero, R. (2013). La evaluación formativa en docencia universitaria y el rendimiento académico del alumnado. *Aula abierta*, 41(2), 23-34. <https://bit.ly/2x6wwKU>
- Gallardo-Fuentes, F., López-Pastor, V. M. & Carter-Thuillier, B. (2018). Efectos de la Aplicación de un Sistema de Evaluación Formativa en la Autopercepción de Competencias Adquiridas en Formación Inicial del Profesorado. *Estudios Pedagógicos*, 44(2), 55-77. <https://doi.org/10.4067/s0718-07052018000200055>
- Hortigüela-Alcalá, D., Abella, V., Delgado, V. & Austin, V. (2016). Influencia del sistema de evaluación empleado en la percepción del alumno sobre su aprendizaje y las competencias docentes. *Revista Infancia, Educación y Aprendizaje*, 2(1), 20-42. <https://doi.org/10.22370/ieya.2016.2.1.582>
- Hortigüela-Alcalá, D., Palacios, A. & López-Pastor, V. M. (2019). The impact of formative and shared or co-assessment on the acquisition of transversal competences in higher education. *Assessment & Evaluation in Higher Education*, 44(6), 933-945. <https://doi.org/10.1080/02602938.2018.150341>
- Hortigüela-Alcalá, D., Pérez-Pueyo, A. & Abella, V. (2015). Perspectiva del alumnado sobre la evaluación tradicional y la evaluación formativa. Contraste de grupos en las mismas asignaturas. *REICE. Revista Iberoamericana sobre Calidad y Cambio en Educación*, 13(1), 35-48. https://www.uv.es/RELIEVE/v21n1/RELIEVEv21n1_ME6.pdf
- López-Pastor, V. M. (Coord.). (2009). *La Evaluación Formativa y Compartida en Educación Superior: propuestas, técnicas, instrumentos y experiencias*. Narcea.
- López-Pastor, V. M. (2012). Evaluación Formativa y Compartida en la Universidad: clarificación de conceptos y propuestas de intervención desde la Red Interuniversitaria de Evaluación Formativa. *Psychology, Society & Education*, 4(1), 117-130. <https://bit.ly/3cVCBud>
- López-Pastor, V. M. (2017). Capítulo 1. Evaluación Formativa y Compartida: Evaluar para aprender y la implicación del alumnado en los procesos de evaluación y aprendizaje. En V. M. López-Pastor & A. Pérez-Pueyo, *Buenas prácticas docentes. Evaluación Formativa y Compartida en Educación: Experiencias de éxito en todas las etapas educativas*, pp. 34-68. León: Universidad de León, secretariado de publicaciones. <https://bit.ly/2IRLwit>
- López-Pastor, V. M., Castejón, J. & Pérez-Pueyo, A. (2012). ¿Implicar al alumnado en la evaluación en la formación inicial del profesorado? Un estudio de caso de evaluación entre iguales de un examen. *Multidisciplinary Journal of Educational Research*, 2(2), 177-201. <https://doi.org/10.4471/remie/2012.09>
- López-Pastor, V. M., Hamodi, C. & López, A. T. (2017). Evaluación Formativa y Compartida en Educación Superior: Buenas Prácticas y Experiencias desarrolladas por la Red de Evaluación Formativa. En C. Eparza (Coord.), *Buenas prácticas en evaluación de aprendizajes en Educación Superior* (pp.13-43). Universidad de los Lagos.
- López-Pastor, V. M., Molina, M., Pascual, C. & Manrique, J. C. (2020). La importancia de utilizar la Evaluación Formativa y Compartida en la Formación Inicial del Profesorado de Educación Física: los Proyectos de Aprendizaje Tutorado como ejemplo de buena práctica. *Retos*, 37, 620-627. <https://bit.ly/2Qhf61A>
- López-Pastor, V. M. & Sicilia-Camacho, A. (2017). Formative and shared assessment in higher education. Lessons learned and challenges for the future. *Assessment & Evaluation in Higher Education*, 42(1), 77-97. <https://dx.doi.org/10.1080/02602938.2015.1083535>
- Lozano, C., Casado, O. M. & Pérez-Pueyo, A. (2017). Capítulo 15. La preparación de exámenes a través de la evaluación formativa y la redacción colectiva de preguntas y respuestas. En V. M. López-Pastor, & A. Pérez-Pueyo, *Buenas prácticas docentes. Evaluación Formativa y Compartida en Educación: Experiencias de éxito en todas las etapas educativas*, pp. 272-285. Universidad de León, secretariado de publicaciones. <https://bit.ly/2IRLwit>
- Martínez-Mínguez, L. (2016). Proyectos de Aprendizaje Tutorados y autoevaluación de competencias profesionales en la formación inicial del profesorado. *Retos*, 29, 242-250. <https://bit.ly/3a2klle>
- Martos, D., Torrent, G., Durbá, V., Saíz, & Tamarit, E. (2014). El desarrollo de la autonomía y la responsabilidad en educación física: un estudio de caso colaborativo en secundaria. *Retos: nuevas tendencias en educación física, deporte y recreación*, 26, 3-8. <https://doi.org/10.47197/retos.v0i26.34386>

- Molina Soria, M., López-Pastor, V. M., Pascual Arias, C. & Barba Martín, R. A. (2019). Los proyectos de aprendizaje tutorado como buena práctica en educación física en primer curso de doble titulación (educación infantil y educación primaria). *Revista de Innovación y Buenas Prácticas Docentes*, 8, 1-14. <https://doi.org/10.21071/ripadoc.v8i1.11988>
- Moss, C. M. & Brookhart, S. M. (2019). *Advancing Formative Assessment in Every Classroom. A guide for Instructional Leaders*. ASCD.
- Nicol, D. J. & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218. <https://doi.org/10.1080/03075070600572090>
- Pérez-Pueyo, A., López-Pastor, V. M., Hortigüela-Alcalá, D. & Gutiérrez, C. (2017). Capítulo 2. Aclaración de los términos implicados en el proceso de evaluación educativa. En V. M. López-Pastor, y A. Pérez-Pueyo, *Buenas prácticas docentes. Evaluación Formativa y Compartida en Educación: Experiencias de éxito en todas las etapas educativas*, pp. 70-91. Universidad de León, secretariado de publicaciones. Recuperado de: <https://bit.ly/2IRLwit>
- Pérez-Pueyo, A., Martínez-Mínguez, L. & Santos, M. (2014). Tres Buenas Prácticas con Evaluación Formativa en Docencia Universitaria. *Revista de Educación Física*, 30(1), 1-15. <https://doi.org/10.1080/13573322.2012.713860.22>
- Romero-Martín, M. R., Asún, S. & Chivité, M. T. (2016). La autoevaluación en expresión corporal en formación inicial del profesorado de educación física: un ejemplo de buena práctica. *Retos. Nuevas tendencias en Educación Física, Deportes y Recreación*, 29, 236-241
- Silva, I. & López-Pastor, V. M. (2015). ¿Cómo vive el alumnado la evaluación en formación inicial del profesorado? @tic revista de innovación educativa, 14. <https://doi.org/10.7203/attic.14.4171>
- Simons, H. (2011). *El estudio de caso: teoría y práctica*. Morata.
- Stake, R. E. (2005). Qualitative case studies. En N. Denzin, y Y. Lincoln (Eds.), *The SAGE handbook of qualitative research* (pp.443-446). Sage Publication.