

Confiabilidad de las valoraciones de los estudiantes. Estudio de sus competencias evaluativas

José Luis Gallego Ortega
Antonio Rodríguez Fuentes
 Universidad de Granada (España)

The accuracy of student's evaluations. Study of their evaluative competences

José Luis Gallego Ortega
Antonio Rodríguez Fuentes
 Universidad de Granada (España)

Resumen

Es sabido que la didáctica universitaria está cambiando con la implantación del EEES y sus nuevos escenarios, recursos y estrategias, al reclamar mayores cotas de implicación estudiantil, en tanto que constructores de su propio aprendizaje. Ello no se está reflejando en la evaluación que sigue primando la tradicional heteroevaluación del docente al discente, aunque con menor peso del examen escrito en beneficio del trabajo procesual. Resta combinar este patrón con otra heteroevaluación como la coevaluación; e, incluso, con la autoevaluación. A ello contribuye el estudio cuasi-experimental pretest-postest de la experiencia innovadora que

Abstract

It is well known that university teaching is undergoing a shift with the introduction of the EHEA and its new scenarios, resources and strategies. We are seeing calls for greater levels of student involvement, as individuals who shape their own learning. This is not being reflected in assessment, which still prioritises the traditional hetero-assessment of the student by the lecturer, although written exams carry less weight than coursework. This model needs to be combined with others models such as peer assessment and even self-assessment. The semi-experimental pre-test/post-test study of the innovative experiment presented here contributes towards this. It outlines

se presenta, que indica las bondades de las fórmulas anteriores a la par que sus requerimientos previos. El grupo control no experimentó mejora de su competencia evaluadora ni mostró confiabilidad de sus puntuaciones diversas por su independencia, a diferencia del experimental que fue sometido a una breve capacitación evaluadora. Además de las medidas de tendencia central (medias) y de dispersión (desviación típica), se acudió a la estadística inferencial, concretamente de contraste (Student y ANOVA) y de correlación (Pearson), para verificar la hipótesis de la confiabilidad de las evaluaciones del alumnado. Es necesario abrir el escenario de la evaluación para conseguir mayor justicia evaluadora, que no objetividad, pero es necesario hacerlo de la forma adecuada, con un tratamiento eficaz como el que presenta este estudio. Heteroevaluaciones tradicionales (profesores) y no tradicionales (coevaluaciones) y autoevaluaciones responsables pueden y deben combinarse en la realidad de las aulas universitarias. La Universidad del futuro así lo demanda y los profesores así debemos cubrirla.

Palabras clave: enseñanza superior, evaluación universitaria, evaluación del docente, evaluación del estudiante, autoevaluación, evaluación formativa.

the benefits of the previous formulas together with their prior requirements. The control group did not experience any improvement in their assessment capabilities, nor did their various scores appear to be reliable as a result of their independence. This differs from the experimental group, who were given some basic assessment training. In addition to central trend measures (averages) and dispersion measures (standard deviation), inferential statistics were also used, specifically contrasts (Student and ANOVA) and correlation (Pearson) to verify the hypothesis regarding the reliability of the students' assessments. Different forms of assessment need to be considered in order to achieve a greater level of assessment fairness, not objectivity. However, this needs to be done adequately and by using an efficient process such as the one presented in this study. The day-to-day reality of university lecture theatres can, and indeed must, include traditional hetero-assessments, group peer assessments and responsible self-assessments. This is what the University of the future demands and we, as lecturers, are the ones who must provide this.

Key words: higher education, university assessment, teacher assessment, student assessment, self-assessment, training assessment.

Introducción

El presente trabajo parte de la premisa justificada por Pérez Juste (2016) sobre el camino que ha de tomar la evaluación en los tiempos actuales para evitar la deriva. Evaluar no es solo calificar sino que es comprobar día a día que las enseñanzas están causando el efecto deseado y los aprendizajes se están produciendo, es decir, que la enseñanza del docente está contribuyendo a la construcción del aprendizaje de los discentes. Cualquier esfuerzo loable sobre esta dimensión alcanza niveles suficientes de justificación en pro de la mejora de la propia realidad educativa, fin último evaluador, y no sólo la calificación del alumnado, objetivo más visible y frecuente (Palacios y López Pastor, 2013). Y ésta

es la justificación que nos mueve en esta ocasión, en un contexto, el universitario, donde conviene, tras la implantación de la convergencia europea de titulaciones y metodologías (EEES), replantear el procedimiento evaluador. Si se trata de pasar del paradigma del “profesor que enseña” al del “alumno que aprende”, se ha de pasar del “profesor que evalúa” al “alumno que colabora”, donde éste deje de ser objeto pasivo de una calificación final, que no siempre corresponde con el avance real en la asignatura, según Moro (2016).

Parte de ese nuevo paradigma docente el hecho de que las enseñanzas no solo corresponden exclusivamente al docente como los aprendizajes tampoco al discente, sino que los roles se intercambian en un nuevo escenario en el cual todos enseñan y todos aprenden. Luego no procede que solo el profesor evalúe, sino que todos pueden hacerlo. No se trata de restar protagonismo profesoral, más al contrario. La evaluación del profesor hacia el alumnado se erige esencial en nuestros actuales sistemas educativos, a diferencia de la evaluación del docente por el alumnado para propiciar ese aprendizaje docente resaltado, aunque sean equiparables en cuanto a oportunidad. Bien realizadas, pueden coexistir y nutrirse en beneficio de su pretendida finalidad de optimización de la realidad. Se concibe pues como una heteroevaluación múltiple frente a la dominante unidireccionalidad tradicional (Gil Flores, 2012), mediante instrumentos típicos de evaluación, principalmente exámenes (López Lozano y Solís Ramírez, 2016). En efecto, Rodríguez Espinosa, Retrespo y Luna (2016) e Ibarra Sáiz y Rodríguez Gómez (2010; 2014) afirman que aunque los estudiantes tienen buena percepción de los procesos evaluativos del docente reclaman fórmulas diferentes a la tradicional como complemento. Porque, según experiencias empíricas realizadas con percepciones del propio alumnado (Lukas, Santiago, Lizasoain y Etxebarria (2017), tal evaluación teórica en modalidad de heteroevaluación final sumativa no refleja lo aprendido ni el esfuerzo del alumno.

Por ello, otras heteroevaluaciones posibles, con fórmulas distintas como la evaluación entre iguales, como la evaluación por pares (Gielen, Docky y Onghena, 2011) o grupales, y la coevaluación compartida dentro del aula, expresiones no sinónimas empero frecuentemente confundidas (Rodríguez Gómez, Ibarra Sáiz y García Jiménez, 2013), cobran sentido en el ámbito del trabajo en grupos masivamente implementado en aulas universitarias tras la apuesta por metodologías activas impulsadas por el EEES (Gallego Noche, Quesada, Gómez y Cubero, 2017; Gámez Montalvo y Torres Martín, 2013; Ibarra Sáiz y Rodríguez Gómez, 2010; 2014). Adicionalmente, mejora la participación de todos en actividades, la retroalimentación de su aprendizaje, la comunicación entre compañeros y la actitud crítica y asunción de ella (Barriopedro, López de Subijana, Gómez Ruano y Rivero, 2016; Gallego Noche, Quesada, Gómez y Cubero, 2017).

La autoevaluación en el marco conceptual de la evaluación como mejora, tanto personal como profesional, debe trazarse como impulso de una actitud y reflexión autocrítica que contribuye a la madurez del estudiantado, personal y profesional (López, 2005). Si requiere cierta base madurativa, el ambiente universitario se erige escenario ideal donde iniciar y propiciar su empleo (Ibarra Sáiz y Rodríguez Gómez, 2007).

Previa a la combinación armónica y estructurada de las anteriores modalidades evaluativas en un modelo secuencial continuado y combinado de evaluación universitaria, ha de probarse la independencia y rigurosidad de sus fórmulas y su combinación. No

se trata de erradicar la subjetividad evaluadora, pretensión inviable, sino que se ajuste a parámetros académicos-pedagógicos y no personales-emocionales. En definitiva, se pretende escudriñar las siguientes dimensiones que se presentan como objetivos operativos de este estudio:

1. Analizar la competencia predictiva del alumnado sobre sus pruebas escritas y en el desarrollo de las prácticas de aula.
2. Descubrir la capacidad mostrada para la coevaluación de compañeros de su grupo propio de trabajo así como del resto de grupos del aula.
3. Cuantificar la confiabilidad e independencia de las evaluaciones realizadas por el alumno al docente, respecto de su propia autoevaluación del examen y prácticas.
4. Estudiar todo lo anterior con alumnos de dos cursos de Magisterio (2º y 4º), con 2 grupos (control y experimental), para contrastar la evolución por cursos (intermedios) y la eficacia del tratamiento.

Método

En consonancia con el objetivo último de investigación que cierra el apartado previo, se empleó un diseño cuasi-experimental con grupos de control (GC) y otros experimentales (GE) sometidos a tratamiento (sobre concienciación de la importancia de la evaluación y estrategias para acometerla, así como la experimentación de la misma) de medidas pares (pretest-postest) y de medidas repetidas (6 en total correspondientes con las prácticas de las asignaturas). Para evaluar la confiabilidad de las puntuaciones otorgadas por el alumnado, tanto a ellos mismos como a los compañeros e incluso al propio profesor, se ha complementado la investigación con un estudio relacional sobre coincidencia e independencia de las puntuaciones ofrecidas.

La población está compuesta por estudiantes universitarios del grado de Magisterio de dos asignaturas, Atención a la Diversidad en Educación Primaria "ADEP" y Atención Psicoeducativa a las Necesidades Educativas Especiales "APNEE" de sendos cursos (2º y 4º) y sendas promociones (2015/16 y 2016/17). En 2º curso de los 4 grupos arbitrariamente elegidos de la asignatura ADEP (C y D de los cursos 2015/16 y 2016/17) hay un total de 289 (155 del GC: curso 2015/16 y 134 del GE: curso 2016/17), de tal manera que la muestra representativa¹ asciende a 166 (83 del GC y 83 del GE). En 4º curso en los 4 grupos aleatoriamente elegidos de la asignatura APNEE (A y B de los cursos 2015/16 y 2016/17) hay un total de 110 (61 del GC: curso 2015/16 y 49 del GE: curso 2016/17), por tanto el tamaño muestral ha de ser de 86 (44 del GC y 44 del GE). La muestra total asciende a 252 alumnos de Magisterio de 2º y 4º cursos.

En cuanto a las variables, son 3 las que se tratan de medir con el alumnado de las asignaturas y los cursos anteriormente descritos:

- Competencia para la autoevaluación y confiabilidad de la misma: por una lado, para la autoevaluación de sus propios exámenes (en su modalidad dual de prueba objetiva y de preguntas de desarrollo); y, por otro, para sus trabajos de aula en

¹ Porcentaje de error: 5%.

el seno de un grupo operativo (5 alumnos, aproximadamente). Concretamente, se mide si existe la deseada relación entre las puntuaciones estimadas por el alumnado y las asignadas posteriormente por los docentes. Del acercamiento de ellas dependerá la confiabilidad de la autoevaluación del alumnado objeto de estudio.

- Competencia para la evaluación de sus compañeros y confiabilidad de la misma: tanto de sus compañeros de grupo de trabajo operativo en su contribución en la elaboración de trabajos prácticos correspondientes a cada unidad didáctica de la asignatura y su exposición en clase, como del resto de compañeros pertenecientes a otros grupos del aula, por sus exposiciones en clase. Concretamente, la independencia de unas puntuaciones con otras. Independencia que responde a la realidad de que cada alumno muestra un esfuerzo particular y diferente al resto, y que evidenciaría la confiabilidad de la coevaluación.
- Competencia para la evaluación del docente y confiabilidad de la misma: tanto para un docente como para otro, ambos responsables de cada una de las asignaturas y de cada uno de los cursos implicados en la investigación. Concretamente, se quiere cuantificar la independencia de una puntuación a uno respecto del otro, y además la independencia de la nota apuntada al profesor con respecto a la nota que cree que el profesor la va a poner a él, puntuaciones que de forma natural deberían ser independientes, y de su verificación, como en los casos anteriores, dependerá la confiabilidad de la heteroevaluación del alumnado.

De otra parte, las variables independientes tomadas en consideración por su interés para el estudio son:

- Curso académico, en sus valores de 2º curso (2º semestre) y 4º curso (primer semestre) del grado de magisterio en Educación Primaria. Se ha optado por ellos porque 2º curso-2º semestre se sitúa como el ecuador de la formación universitaria, y 4º curso-1º semestre en el culmen del aprendizaje de la carrera.
- Tratamiento recibido por el alumnado para desarrollar esta tarea evaluadora, distinguiendo entre un grupo denominado control que no recibió este tratamiento y otro experimental que sí. El tratamiento consiste en 5 acciones resumidas mediante el acrónimo CREER:
 - C → Concienciación sobre la importancia de la evaluación mediante explicación y reflexión compartida.
 - R → Rúbricas de evaluación (autoevaluación y coevaluación) entregadas a cada uno y compromiso de empleo ético (Saiz y Bol, 2014).
 - E → Experimentación-ensayo de evaluación de sí mismos y de otros compañeros y profesor.
 - E → Establecer unos tiempos específicos dentro de cada práctica para cumplimentar las evaluaciones correspondientes.
 - R → Regulación y supervisión del docente de la cumplimentación de la evaluación continua de las prácticas de clase (Moro, 2016).

Para la recogida de datos sobre previsión de puntuaciones se incluyó una tabla en el encabezado de pruebas que demandaba al alumnado puntuaciones estimadas en pruebas objetivas y de preguntas a desarrollar. Cuando revisaban sus exámenes habían de completar en espacio reseñado para ello la calificación que otorgaban al docente. De esta forma se completaron 2 de las dimensiones requeridas para el estudio: estimaciones de sus evaluaciones o autoevaluación de pruebas escritas (que se presentarán en el subapartado primero de resultados) y valoración del docente o heteroevaluación (subapartado tercero de resultados).

Durante las prácticas de asignatura se entregó una ficha personal a cada alumno con la indicación de que para cada una de las 6 prácticas, correspondientes con cada una de los 6 Unidades Didácticas, cumplimentaran, en su momento de exposición, la puntuación que se asignaban a ellos mismos y al resto de sus compañeros de su grupo de trabajo (Cfr. Tabla 1). En el reverso de la ficha se demandaba, también para cada práctica, las puntuaciones correspondientes a cada uno del resto de los grupos del seminario (Cfr. Tabla 2). Al GE esta ficha se le entregaba acompañada de otra a modo de rúbrica de evaluación y se hacía un ensayo previo, además era supervisada por el docente. De esta forma se recogían los datos que se presentan los apartados siguientes sobre autoevaluación de sus prácticas (subapartado primero de resultados) y coevaluación de sus compañeros de grupo de trabajo y del resto de grupos (subapartado segundo de resultados).

Tabla 1. Anverso de la ficha de evaluación intragrupo.

Alumno:			Curso:		Seminario:		Grupo:	
(0 a 10)	Autoevaluación		Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	\bar{x}
Práctica 1								
Práctica 2								
Práctica 3								
Práctica 4								
Práctica 5								
Práctica 6								
\bar{x}								

Tabla 2. Reverso de la ficha de evaluación intergrupo.

Alumno:			Curso:		Seminario:		Grupo:	
(0 a 10)	Grupo propio		Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	\bar{x}
Práctica 1								
Práctica 2								
Práctica 3								
Práctica 4								
Práctica 5								
Práctica 6								
\bar{x}								

En cuanto al análisis de datos, además de las medidas de tendencia central (\bar{x}) y dispersión de datos (s), se recurrió a la estadística inferencial para detectar y cuantificar tanto las diferencias entre grupos, especialmente entre GC y GE, acudiendo al cálculo de la t de Student (t), como las correlaciones ente puntuaciones, acudiendo en este caso al cálculo de rho de Pearson (r). En ocasiones puntuales, para el cálculo de las diferencias entre pruebas (más de dos), el cálculo de la ANOVA (F) ha sustituido al de la t de Student. Se justifica el empleo de estas pruebas paramétricas por el tamaño de los grupos ($n > 30$) y la distribución homogénea de datos (homocedasticidad).

Resultados

Interesan al estudio tres tipos de competencias discentes para la evaluación universitaria. Para cuantificar su propia actuación, en exámenes y otras medidas; para valorar la de sus compañeros de grupo y aula, y para analizar el esfuerzo y capacidad del docente. Al tratarse de habilidades diferentes se presentan por separado.

Confiabilidad de la autoevaluación

Conciérne a la competencia auto-evaluadora del alumnado universitario, tanto sobre sus pruebas escritas a modo de exámenes (tipo pruebas objetivos y de desarrollo) como de su trabajo desarrollado en las prácticas de clase durante el desarrollo de la asignatura. La primera se contrasta con la evaluación otorgado por el docente en dichas pruebas; la segunda con las notas otorgadas por el resto de agentes del acto didáctico: el profesor responsable de la asignatura, sus compañeros del grupo de trabajo operativo y el resto de compañeros del resto de grupos del seminario de la asignatura.

A) Para las pruebas escritas: exámenes. En este caso se trata de valorar la capacidad predictora del alumno, contrastando las puntuaciones estimadas por cada alumno en pruebas de distinta naturaleza (una objetiva y otra de preguntas) con las puntuaciones otorgadas por el docente de la asignatura. A continuación, se observan los resultados de los análisis de correlación entre notas estimadas y reales (r de Pearson) y las diferencias entre los grupos control (GC) y experimental (GE) mediante inferenciales (t de Student), y sus comentarios correspondientes. Se han excluido de la tablas los resultados de las pruebas escritas iniciales de las asignaturas, consideradas como medidas pretest, en las que las diferencias entre medias ($t=3.407$) no alcanzan a ser significativas ($p=0.523$) al contrastar los GC tanto de 2º curso: prueba objetiva ($\bar{x}=7.525$; $s=2.226$) y de desarrollo ($\bar{x}=6.025$; $s=3.004$) como de 4º curso ($\bar{x}=8.300$; $s=2.852$ para prueba objetiva y $\bar{x}=7.143$; $s=2.901$ para la otra) con los GE tanto de 2º ($\bar{x}=7.490$; $s=2.422$ en objetiva y $\bar{x}=5.893$; $s=3.411$ en desarrollo) como de 4º (prueba objetiva: $\bar{x}=7.750$; $s=2.548$ y de desarrollo: $\bar{x}=7.750$; $s=2.548$). Dada la ausencia de diferencias en esta medición inicial cobran valor las diferencias significativas ($p > 0.05$) en la evaluación final que se visualizan a continuación (Cfr. Tabla 3), empleando los mismos estadísticos de contraste (t de Student) y de relación (r de Pearson).

Tabla 3. Relación entre notas estimadas (alumnado) y reales (docente) en pruebas de evaluación.

Nota Examen	Curso 2º				Curso 4º			
	GC		GE		GC		GE	
	Estimada	Real	Estimada	Real	Estimada	Real	Estimada	Real
Prueba Objetiva (test)	$\bar{x}=6.497$ ($s=1.343$)	$\bar{x}=4.343$ ($s=2.427$)	$\bar{x}=6.193$ ($s=1.814$)	$\bar{x}=5.918$ ($s=1.961$)	$\bar{x}=7.239$ ($s=2.720$)	$\bar{x}=6.886$ ($s=2.343$)	$\bar{x}=6.693$ ($s=2.027$)	$\bar{x}=6.438$ ($s=2.057$)
	r=0.018 (p=0.095)		r=0.890 (p=0.000)		r=0.471 (p=0.001)		r=0.958 (p=0.000)	
	t=1.228 (p=0.000)				t=-2.550 (p=0.013)			
Prueba de Preguntas (cortas)	$\bar{x}=5.650$ ($s=1.548$)	$\bar{x}=3.791$ ($s=4.119$)	$\bar{x}=5.621$ ($s=1.778$)	$\bar{x}=4.631$ ($s=1.963$)	$\bar{x}=7.193$ ($s=1.782$)	$\bar{x}=6.612$ ($s=2.256$)	$\bar{x}=6.966$ ($s=1.930$)	$\bar{x}=6.653$ ($s=1.962$)
	r=0.153 (p=0.168)		r=0.832 (p=0.000)		r=0.305 (p=0.044)		r=0.927 (p=0.000)	
	t=-1.678 (p=0.045)				t=0.574 (p=0.036)			

Respecto del curso 2º los datos apuntan diferencias claras. Si bien tanto el GC como el GE los alumnos se puntúan (estimación) por encima de lo real en ambas pruebas (objetivos y desarrollo) lo cierto es que en el caso del GE las diferencias no resultan significativas y sí la correlación mediante Pearson (que además resulta alta: $r=0.890$ ($p=0.000$) para la prueba objetiva y $r=0.832$ ($p=0.000$) para la prueba de preguntas a desarrollar), a diferencia del GC, como indica la prueba de Student, tanto en prueba objetiva ($t=1.228$; $p=0.000$) como de desarrollo ($t=-1.678$; $p=0.045$). La interpretación que cabe hacer de estos datos es que, partiendo de que no existían diferencias entre GC y GE en su inicio (pretest), lo cierto es que en las medidas posttest (postratamiento) el GC no realiza una buena predicción de sus notas en las distintas pruebas mientras que el GE sí lo hace, debido al "tratamiento" recibido; por tanto mejora su capacidad autoevaluatora.

Los datos posttest se comportan de partida de manera diferente en el caso de 4º curso aunque culminan igual. Esto es, las puntuaciones estimadas suelen ser en ambos grupos (GC y GE) mayores que las reales, aunque no tanto como en el curso anterior, sino que son más similares entre ellas; de hecho, en ambos grupos correlacionan positivamente, aunque en el GE mucho más intensamente ($r=0.958$ ($p=0.000$) en la prueba objetiva, y $r=0.927$ ($p=0.000$) en la de desarrollo) que en el GC ($r=0.471$ ($p=0.001$) en objetiva y $r=0.305$ ($p=0.044$) en la de desarrollo). De tal manera que las diferencias entre ambos grupos también resultan significativas ($t=0.574$; $p=0.036$). Luego si bien los alumnos de 4º demuestran una moderada predicción de sus puntuaciones *per se* (quizá como fruto de su madurez profesional en el culmen de su formación docente), esta capacidad mejora con la toma de conciencia y experimentación del tratamiento aplicado al GE.

B) Para las prácticas de asignatura. Se exponen, en esta ocasión, los cálculos estadísticos realizados en la tabla siguiente y los comentarios a continuación sobre la supuesta relación (similitud) que cabría esperar entre la evaluación no escrita de sí mismo (autoevaluación) respecto de otras evaluaciones recibidas de otras personas (del profesor, de su grupo de trabajo y del resto de grupos de clase) hacia cada alumno. Los resultados de partida, obtenidos en la valoración de la primera práctica (del total de 6), han sido tomados como evaluación pretest, y cuyos resultados son similares según la ausencia de diferencias significativas entre ellos ($t=2.575$; $p=0.159$). A partir de estos

resultados iniciales, se han tomado las 5 evaluaciones restantes en cada grupo. En el GC de ambos cursos, por separado, no se han encontrado diferencias significativas de una a otra práctica, de acuerdo con el cálculo de la t de Student ($p > 0.05$ en todos los casos) ni en todas en su conjunto, según el nivel de significatividad ($p = 0.358$) arrojado por el cálculo de la ANOVA de un factor ($F = 12.379$). Idénticos cálculos estadísticos, sin embargo, han puesto de manifiesto diferencias en el avance de la autoevaluación de unas prácticas a otras en el GE ($F = 9.562$, $p = 0.007$), siendo cada práctica acompañada con la introducción de una herramienta nueva de las 5 que conforman el tratamiento que aparecen explícitas en el apartado anterior. Éste resultado hace interesante la presentación en la tabla siguiente (4) de los resultados de la última práctica considerado posttest pues coinciden con la culminación del tratamiento, y sus diferencias entre GC y GE.

Tabla 4. Relación entre notas estimadas (alumnado) y recibida por otros.

	Curso 2º		Curso 4º	
	GC	GE	GC	GE
Auto-evaluación \bar{x}	$\bar{x} = 8.356$ ($s = 0.461$)	$\bar{x} = 7.531$ ($s = 0.615$)	$\bar{x} = 8.591$ ($s = 0.807$)	$\bar{x} = 7.558$ ($s = 0.829$)
	$t = 2.081$ ($p = 0.039$)		$t = -0.189$ ($p = 0.030$)	
Nota \bar{x} otorgada por el docente	$\bar{x} = 7.15$ ($s = 2.596$) $r = 0.117$ ($p = 0.07$)	$\bar{x} = 7.250$ ($s = 3.507$) $r = 0.471$ ($p = 0.015$)	$\bar{x} = 7.371$ ($s = 3.219$) $r = 0.473$ ($p = 0.041$)	$\bar{x} = 7.420$ ($s = 3.014$) $r = 0.595$ ($p = 0.002$)
	$t = -2.764$ ($p = 0.45$)		$t = 2.356$ ($p = 0.842$)	
Nota \bar{x} otorgada por su grupo	$\bar{x} = 8.241$ ($s = 1.809$) $r = 0.635$ ($p = 0.061$)	$\bar{x} = 7.321$ ($s = 3.978$) $r = 0.659$ ($p = 0.037$)	$\bar{x} = 7.807$ ($s = 1.561$) $r = 0.593$ ($p = 0.026$)	$\bar{x} = 7.250$ ($s = 1.720$) $r = 0.854$ ($p = 0.000$)
	$t = -3.405$ ($p = 0.042$)		$t = 1.859$ ($p = 0.002$)	
Nota \bar{x} otorgada por otros grupos	$\bar{x} = 7.732$ ($s = 1.223$) $r = 0.304$ ($p = 0.005$)	$\bar{x} = 7.251$ ($s = 1.302$) $r = 0.592$ ($p = 0.019$)	$\bar{x} = 7.552$ ($s = 1.340$) $r = 0.432$ ($p = 0.003$)	$\bar{x} = 7.115$ ($s = 1.621$) $r = 0.723$ ($p = 0.007$)
	$t = -1.221$ ($p = 0.006$)		$t = 2.079$ ($p = 0.025$)	

Los datos sobre la autoevaluación realizada por el alumno sobre sus prácticas de aula refuerzan los resultados anteriores. Los alumnos de 2º curso se evalúan a sí mismos por encima que a sus compañeros y al profesor, lo cual acontece tanto en las medias pretest como posttest; no obstante en las posttest se aprecian ciertas diferencias entre GC y GE. Como en las pretest, resulta sospechoso que las mayores puntuaciones posttest y, por tanto, la mayor correlación se produce entre esas generosas puntuaciones propias y las no menos generosas puntuaciones asignadas a sus compañeros de grupo y la menor entre autoevaluación y evaluación del docente (Cfr. Tabla 4). No obstante, en el caso del GE las correlaciones entre las puntuaciones de autoevaluación y las demás son más similares que las del GC (Cfr. Tabla 4), correlacionando en todos los casos las puntuaciones sobre el esfuerzo del alumno; de tal manera que las diferencias entre GC y GE resultan significativas, como indican los datos estadísticos contenidos en la tabla anterior, salvo en el caso de evaluaciones propias y del docente, donde el nivel de significación ($t = -2.764$; $p = 0.45$) indica lo contrario.

El mismo patrón que en la autoevaluación de la prueba objetiva se dibuja también en esta ocasión acerca del examen de preguntas en el caso del curso 4º. El escenario mejora en contraste con el curso anterior dada la correlación de la anterior

autoevaluación respecto de la otorgada por el docente, los miembros del grupo de trabajo y el resto de miembros de otros grupos, observándose cierta coincidencia moderada entre puntuaciones en el GC, coincidencia que se hace más intensa en el GE (Cfr. Tabla 2). Ello hace que las diferencias entre GC y CE resulten significativas, salvo para el caso de las puntuaciones otorgadas por el profesorado ($t=2.356$; $p=0.842$) (Cfr. Tabla 2).

Confiabilidad de la coevaluación

En esta ocasión, se valora la competencia que muestra el alumnado para la evaluación de iguales:

A) Compañeros directos de su grupo de trabajo. Como quiera que no son significativas ($p=0.487$) de partida (pretest) las diferencias entre las coevaluaciones realizadas por los alumnos del GC y GE ($t=4.575$) a sus compañeros de grupos de trabajo y sí en la repeticiones de evaluaciones de las prácticas restantes, en su conjunto, para el caso del GE ($F=5.780$, $p=0.050$), cabe presentar la tabla 3 con las evaluaciones finales de la última práctica (postest), donde se plasman diferencias entre GC y GE, mediante *t* de Student, tanto para 2º ($t=-3.681$; $p=0.000$) como para 4º ($t=4.232$; $p=0.000$). También puede apreciarse la supuesta independencia que debiera existir entre las evaluaciones realizadas para sí mismos (autoevaluaciones) y sus compañeros de grupo de trabajo, así como entre la evaluación de sus propios compañeros de su grupo (intragrupos) y otros compañeros de otros grupos (intergrupos) en sendos GE, mediante rho de Pearson.

Tabla 3. Relación entre coevaluación del grupo propio de prácticas y otras coevaluaciones y de sí mismos.

	Curso 2º		Curso 4º	
	GC	GE	GC	GE
Evaluación Intragrupo	$\bar{x}=8.051$ ($s=0.443$)	$\bar{x}=8.334$ ($s=0.544$)	$\bar{x}=8.576$ ($s=0.820$)	$\bar{x}=7.768$ ($s=0.965$)
	$t=-3.681$ ($p=0.000$)		$t=4.232$ ($p=0.000$)	
Autoevaluación prácticas de aula	$\bar{x}=8.356$ ($s=0.461$) $r=0.656$ ($p=0.000$)	$\bar{x}=7.531$ ($s=0.615$) $r=0.062$ ($p=0.578$)	$\bar{x}=8.591$ ($s=0.807$) $r=675$ ($p=0.000$)	$\bar{x}=7.558$ ($s=0.829$) $r=0.281$ ($p=0.065$)
Evaluación intergrupos	$\bar{x}=7.652$ ($s=0.447$) $r=0.636$ ($p=0.001$)	$\bar{x}=8.406$ ($s=0.505$) $r=0.367$ ($p=0.005$)	$\bar{x}=8.275$ ($s=0.543$) $r=0.520$ ($p=0.050$)	$\bar{x}=7.650$ ($s=0.879$) $r=0.321$ ($p=0.012$)

Dentro de la coevaluación, la evaluación de los miembros de su grupo de trabajo operativo alcanza las puntuaciones más altas (Cfr. Tabla 3), incluso en el caso de los alumnos del GE por encima de la autoevaluación, en sendos cursos, a diferencia de los alumnos del GC, donde la autoevaluación supera a la del grupo propio, en sendos grupos aunque de manera más acusada en 2º. Tales son las diferencias entre GC y GE que resultan significativas ($t=-3.681$ ($p=0.000$) en 2º y $t=4.232$ ($p=0.000$) en 4º).

En efecto, las evaluación que hace el GC de su grupo de 2º y 4º correlaciona con sus autoevaluaciones ($r=0.656$ ($p=0.000$) para 2º y $r=675$ ($p=0.000$) para 4º), lo cual no respeta la independencia que debieran tener ambas evaluaciones de personas diferentes y pudiera apuntar cierto pacto táctico y tácito para conseguir puntuaciones altas y similares entre todos los miembros de cada grupo de trabajo. Satisfactoriamente,

esto no acontece en ninguno de los GE ($r=0.062$ ($p=0.578$) para 2º y $r=0.281$ ($p=0.065$) para 4º), lo que apunta cierta eficacia de la reflexión y ensayo compartidos sobre la evaluación, coevaluación en este caso.

Independencia que no se ha conseguido al observar la relación entre la evaluación interna de los miembros de su grupo de trabajo respecto de los otros grupos. En estos casos, las puntuaciones de otros grupos diferentes al suyo es siempre menor que la del propio grupo, lo cual no parece responder siempre a la realidad, y existe cierta relación entre unas y otras. No obstante, en los alumnos del GE de sendos grupos la correlación entre las anteriores evaluaciones es baja ($r=0.367$ ($p=0.005$) para 2º y $r=0.321$ ($p=0.012$) para 4º) mientras que en el GC es media ($r=0.636$ ($p=0.001$) para 2º y $r=0.520$ ($p=0.050$) para 4º), lo cual reduce la dependencia existente entre ellos que pese a ello no se ha podido salvar.

B) Otros compañeros de su clase, pertenecientes al resto de grupos de trabajo. Tanto para 2º como para 4º los alumnos de distintos grupos no mostraron divergencias significativas ($p=0.956$ y $p=0.356$, respectivamente) en sus evaluaciones iniciales respecto a su evaluación de otros grupos diferentes a su grupo de trabajo, según el estadístico t de Student ($t=7.569$ y $t=5.455$, respectivamente). De ahí que las diferencias finales entre GC y GE sean relevantes y presentadas en la tabla 6. Máxime cuando la ANOVA entre medidas repetidas entre los 6 grupos (distintas evaluaciones) alcanza significatividad ($F=4.325$; $p=0.003$) en el progreso del GE, a diferencia del GC ($F=9.876$; $p=0.369$).

Tabla 6. Relación entre coevaluación de otros grupos con el suyo propio y consigo mismos.

	Curso 2º		Curso 4º	
	GC	GE	GC	GE
Evaluación intergrupos	$\bar{x}=7.652$ ($s=0.447$)	$\bar{x}=8.406$ ($s=0.505$)	$\bar{x}=8.275$ ($s=0.543$)	$\bar{x}=7.650$ ($s=0.879$)
	$t=-9.988$ ($p=0.000$)		$t=5.962$ ($p=0.000$)	
Autoevaluación prácticas de aula	$\bar{x}=8.356$ ($s=0.461$) $r=0.369$ ($p=0.001$)	$\bar{x}=7.531$ ($s=0.615$) $r=0.104$ ($p=0.348$)	$\bar{x}=8.591$ ($s=0.807$) $r=0.289$ ($p=0.059$)	$\bar{x}=7.558$ ($s=0.829$) $r=0.264$ ($p=0.084$)
Evaluación Intragrupo	$\bar{x}=8.051$ ($s=0.443$) $r=0.636$ ($p=0.001$)	$\bar{x}=8.334$ ($s=0.544$) $r=0.367$ ($p=0.005$)	$\bar{x}=8.576$ ($s=0.820$) $r=0.520$ ($p=0.050$)	$\bar{x}=7.768$ ($s=0.965$) $r=0.321$ ($p=0.012$)

Para los datos de la Tabla 6 sirven los comentarios del punto anterior. Cabe añadir que la evaluación de otros grupos obtiene menores puntuaciones que las demás evaluaciones. Que existen diferencias entre la competencia para ello del GE y GC ($t=-9.988$ ($p=0.000$) para 2º y $t=5.962$ ($p=0.000$) para 4º), a favor de los alumnos del GE que muestran independencia respecto de las autoevaluaciones, apuntada por la inexistente correlación entre ambas ($r=0.104$ ($p=0.348$) para 2º y $r=0.264$ ($p=0.084$) para 4º), a diferencia del alumnado del GC donde la correlación es baja pero existe ($r=0.369$ ($p=0.001$) para 2º y $r=0.289$ ($p=0.059$) para 4º). Respecto de la relación con la evaluación interna de cada uno de sus grupos, cabe insistir en lo que indicado en el punto anterior: solo se ha conseguido reducir la correlación baja del GE ($r=0.367$ ($p=0.005$) para 2º y $r=0.321$ ($p=0.012$) para 4º) respecto de la media del GC ($r=0.636$ ($p=0.001$) para 2º y $r=0.520$ ($p=0.050$) para 4º) que denota cierta dependencia entre ambas.

Confiabilidad de la heteroevaluación

Por último, tanto en la prueba escrita inicial como en la final se les demandó la valoración cuantitativa del docente o heteroevaluación (evaluación de otro con diferente rol al alumno). Tampoco en esta ocasión difirieron las puntuaciones pretest en sendos grupos, GC y GE, de sendos cursos, 2º y 4º, según los valores de significativad ($p > 0.05$) vinculados al valor de t de Student. Contrariamente, en la medida posttest tanto para uno como para otro curso se detectaron diferencias, cuantificadas en la tabla 7, entre grupos.

Tabla 7. Relación entre heteroevaluación al profesor y sensación de evaluaciones propias.

	Curso 2º		Curso 4º	
	GC	GE	GC	GE
Nota puesta a los docentes	$\bar{x}=8.747$ ($s=0.994$)	$\bar{x}=9.149$ ($s=0.766$)	$\bar{x}=8.657$ ($s=1.119$)	$\bar{x}=8.824$ ($s=0.549$)
	$t=-2.917$ ($p=0.004$)		$t=-6.207$ ($p=0.000$)	
Percepción nota examen	$\bar{x}=6.132$ ($s=1.111$) $r=0.880$ ($p=0.000$)	$\bar{x}=5.886$ ($s=1.535$) $r=0.322$ ($p=0.003$)	$\bar{x}=7.108$ ($s=1.467$) $r=0.495$ ($p=0.001$)	$\bar{x}=6.830$ ($s=1.824$) $r=0.233$ ($p=0.027$)
Autoevaluación prácticas	$\bar{x}=8.356$ ($s=1.282$) $r=0.792$ ($p=0.000$)	$\bar{x}=8.531$ ($s=0.615$) $r=0.199$ ($p=0.071$)	$\bar{x}=8.591$ ($s=0.807$) $r=0.188$ ($p=0.021$)	$\bar{x}=8.558$ ($s=0.829$) $r=0.049$ ($p=0.754$)

Se pueden observar en la Tabla 7 las diferencias significativas entre ambos grupos: GC y GE ($t=-2.917$ ($p=0.004$) para 2º y $t=-6.207$ ($p=0.000$) para 4º). Además y en parte causa de ello, se observa que en el GC estas puntuaciones al docente guardan dependencia o correlación alta tanto con la percepción que creen sacar en los exámenes ($r=0.880$ ($p=0.000$) para 2º y $r=0.495$ ($p=0.001$) para 4º) y con la percepción de su calificación en las prácticas ($r=0.792$ ($p=0.000$) para 2º y $r=0.188$ ($p=0.021$) para 4º). Es decir, que evalúan al docente en función de cómo creen que les va en el trascurso de la asignatura. Esto no ocurre en el caso de los alumnos del GE, en los cuales se observa una menor correlación entre la valoración numérica que asignan al docente y la percepción de sus puntuaciones en el examen ($r=0.322$ ($p=0.003$) para 2º y $r=0.233$ ($p=0.027$) para 4º) y en sus prácticas (baja para 2º: $r=0.199$ ($p=0.071$) y nula para 4º: $r=0.049$ ($p=0.754$)). Sin duda este último caso denota mayor desarrollo de la competencia evaluadora.

Discusión y conclusiones

En el presente estudio se ha pretendido analizar la competencia y confiabilidad de las evaluaciones vertidas por el alumnado universitario sobre sí mismo, sobre sus compañeros más cercanos (del grupo de trabajo) y menos (otros grupos) y del docente. De modo general, cabe concluir que los estudiantes universitarios no manifiestan *per se* una buena competencia para la evaluación, como se desprende de:

- La escasa objetividad para la autoevaluación en las pruebas realizadas, si bien los alumnos más reflexivos aprueban en esta dimensión, especialmente en las pruebas objetivas. Saiz y Payo (2012) detectan diferencias significativas entre la autopercepción del estudiante sobre su propio aprendizaje antes y después de la prueba tradicional de evaluación y la calificación asignada por el docente, debido a la sobrevaloración de las actuaciones realizadas por el alumnado (Ibarra

Sáiz y Rodríguez Gómez, 2012). Según estudios como los de Saiz y Marín (2011) y Pascual Gómez, Lorenzo Llamas y Monge (2015) se desprenden diferencias significativas entre las evaluaciones realizadas por uno (profesor) y por otros (alumnos), de ahí que su complementación resulte totalmente pertinente. A estos resultados y con un diseño similar llegan Saiz y Payo (2012), tras una evaluación pre y post-instrucción, aunque solo con autoevaluaciones y con una instrucción diferente, apareciendo correlación entre las puntuaciones post-instrucción y la calificación final del docente, como acontece en este estudio. Lo anterior camina en la misma dirección de lo hallado en esta investigación sobre diferencias entre puntuaciones autoestimadas y puntuaciones asignadas del profesor, tanto para pruebas objetivas (tipo test) como no objetivas (preguntas cortas).

- b. La claramente excesiva generosidad en sus autovaloraciones del trabajo práctico, “infladas” como descubrieran Wheeler, Landan y Dunleay (2005) o “sobrevaloradas” como descubrieran Pascual Gómez, Lorenzo Llamas y Monge (2015), así como también se extiende tal generosidad a las valoraciones de compañeros de su propio grupo y, algo menos, en las de otros grupos, siendo mayor la sobrevaloración en el caso de alumnos de 2º curso, por tanto más deficiente por cuanto menos ajustado a la realidad. Se han evidenciado ciertos problemas relacionados con la coevaluación, pero sin duda igual que concluyeran Zaragoza, Luis y Manrique (2009) y Pascual Gómez, Lorenzo Llamas y Monge (2015) son más las oportunidades y ventajas que los inconvenientes.
- c. La considerable dependencia entre puntuaciones asignadas al docente y la calificación en la asignatura, esto es, las puntuaciones al docente depende estrechamente de la puntuación que este asigna al examen del estudiante, correlación que nuevamente es mayor en los alumnos de 2º que en los de 4º curso. Esto es interpretado como negativo, puesto que deberían de tratarse de puntuaciones independientes. Mismos datos que los obtenidos por Sluijsmans, Dochy y Moerkerke (1999).

Tras concluir de modo similar, Wheeler, Landan y Dunleay (2005) responsabilizan a la falta de formación del estudiante para tal labor, también apoyada por Ibarra Sáiz y Rodríguez Gómez (2012). Después del programa CREER los resultados han mejorado sustancialmente, lo cual coincide en parte con los estudios presentados por Rodríguez Espinosa, Retrepo y Luna (2016), quienes descubren que en una segunda fase de autoevaluaciones los resultados mejoran considerablemente; luego a la falta de formación cabría añadir falta de experiencia.

Sin embargo, aparecen más divergencias con lo apuntado por Gessa (2011) respecto a la similitud entre las evaluaciones realizadas por el docente y el alumnado, que le sirve para probar la buena competencia evaluadora; aunque lo cierto es que a medida que se avanza en el curso de los participantes de la presente investigación, los datos evaluativos dibujan un perfil más parecido al obtenido por la investigadora. En efecto, se ha observado una tímida mejoría tanto en la autoevaluación, coevaluación y heteroevaluación del alumnado del último curso respecto del curso intermedio, pero aun así siguen observándose deficiencias a modo de falta de competencia mostrada o falta de independencias de unas evaluaciones a otras y de unos evaluados a otros. Sí se logran, sin embargo, niveles aceptables de competencia y confiabilidad tras seguir un

simple tratamiento de eficacia probada como el usado en esta experiencia (CREER) con los alumnos. Luego se abre un espacio para la esperanza de la combinación de diferentes modalidades y agentes evaluadores, lo que ya apuntaba García Jiménez (2015).

De tal suerte que se desprende una proyección bifronte de la experiencia: a) que se puede acometer este sistema para hacer una evaluación más variada y completa, no tanto más objetiva pues ahí ni entró la investigación ni interesa a los investigadores, convencidos de que nunca lo será absolutamente; b) que toda vez que funciona, de la actuación del alumno en la evaluación propia y también de los demás se logra la deseada autorregulación y mejora del propio aprendizaje (Mingorance y Calvo, 2013), dado el feedback que la autoevaluación y la coevaluación suponen, aspecto que también resalta Gessa (2011). Por extensión de una y otra se mejora la propia implicación personal y en los trabajos en grupo, según descubrió Barriopedro *et al.* (2016) se avanza hacia una evaluación más actual, que deje de concebirse como el punto de llegada para pasar a ser el de partida (Gessa, 2011). Como prospectiva, además de continuar y perpetuar esta línea de investigación con otros diseños y enfoques de investigación (por ej. más cualitativos), se propone el empleo de una plataforma virtual que se está gestando para facilitar el compendio de todas las puntuaciones de todos los implicados, de manera que el programas con las ponderaciones requeridas al docente, como responsable último de la calificación y supervisor de la plataforma, calcule la puntuación global única. Lo innovador es la propia plataforma, que no el medio utilizado, dado que otros autores (Carrizosa y Gallardo Ballester, 2015; García Jiménez, 2015) han propuesto el empleo de herramientas tecnológicas, como formularios google docs para acometer la evaluación continua. A este respecto, hay que apuntar que en el estudio de Quesada, Rodríguez Gómez e Ibarra Sáiz (2017) la utilización de medios electrónicos para la evaluación supone la mayor deficiencia del profesorado, junto al diseño de instrumentos de evaluación, luego quizás se requiera de ciertas acciones formativas.

Cabe añadir que el grado de implicación y participación del alumnado ha sido muy adecuado a la importancia de la experiencia. Ya este hecho fue adelantado por Gessa (2011) y Rodríguez Espinosa, Retrepo y Luna (2016). Y han reconocido, adicionalmente, como era objetivo colateral de este proyecto, el valor añadido del esfuerzo del mismo en aras a su desarrollo evaluativo para el ejercicio docente inminente (primero en práctica y después en su incorporación laboral).

Ahora bien, sin perjuicio de la conveniencia de acometer medidas como las anteriores en beneficio de la contextualización apropiada del desarrollo de esta competencia, no resulta menos meritorio continuar esta línea de investigación. Por ejemplo, mediante estudios longitudinales, disponiendo de los mismos participantes a lo largo de toda su formación universitaria. Ampliando los cursos, es decir, incluyendo a estudiantes de primer y tercer Grado, para dibujar la evolución precisa producida, así como el punto de partida. Ampliando alumnado perteneciente a diferentes grupos para controlar en esta ocasión variables intervinientes como consecuencia del profesorado responsable de su formación. Ídem con otras Facultades y Universidades, con diferentes programas y recursos. Incluso variando los instrumentos de recogida de datos y herramientas de análisis de los mismos, como por ejemplo la entrevista o análisis de tareas y/o errores, así como el proceso seguido, por ejemplo, adelantar la valoración sobre el docente a antes de conocer la calificación que éste emite finalmente sobre la asignatura.

Referencias bibliográficas

- Barriopedro, M., López de Subijana, C., Gómez Ruano, M. A. y Rivero, A. (2016). La coevaluación como estrategia para mejorar la dinámica del trabajo en grupo: una experiencia en Ciencias del Deporte. *Revista Complutense de Educación*, 27(2), 571-584. https://doi.org/10.5209/rev_RCED.2016.v27.n2.46811
- Carrizosa, A. y Gallardo Ballester, J. I. (2015). Autoevaluación, Coevaluación y Evaluación de los aprendizajes. Ponencia presentada en *III Jornadas sobre docencia del Derecho y TICs*.
- Gallego Noche, B., Quesada, M. A., Gómez Ruiz, J. y Cubero, J. (2017). La evaluación y retroalimentación electrónica entre iguales para la autorregulación y el aprendizaje estratégico en la universidad: la percepción del alumnado. *REDU. Revista de Docencia Universitaria*, 15(1), 127-146. <https://doi.org/10.4995/redu.2017.5991>
- Gámez Montalvo, M. J. y Torres Martín, C. (2013). Las técnicas de grupo como estrategia metodológica en la adquisición de la competencia de trabajo en equipo de los alumnos universitarios. *Journal for Educators, Teachers and Trainers*, 4(1), 14-25.
- García Jiménez, E. (2015). La evaluación del aprendizaje: de la retroalimentación a la autorregulación. El papel de las tecnologías. *RELIEVE*, 21(2), 1-24. <https://doi.org/10.7203/relieve.21.2.7546>
- Gessa, A. (2011). La coevaluación como metodología complementaria de la evaluación del aprendizaje. Análisis y reflexión en las aulas universitarias. *Revista de Educación*, 345, 749-763.
- Gielen, S., Docky, F. y Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education*, 36(2), 137-155. <https://doi.org/10.1080/02602930903221444>
- Gil Flores, J. (2012). La evaluación del aprendizaje en la universidad según la experiencia de los estudiantes. *Estudios sobre Educación*, 22, 133-153
- Ibarra Sáiz, M^a. S. y Rodríguez Gómez, G. (2007). El trabajo colaborativo en las aulas universitarias. Reflexiones desde la autoevaluación. *Revista de Educación*, 344, 355-375.
- Ibarra Sáiz, M^a. S. y Rodríguez Gómez, G. (2010). Aproximación al discurso dominante sobre la evaluación del aprendizaje en la universidad. *Revista de Educación*, 351, 385-407.
- Ibarra Sáiz, M^a. S. y Rodríguez Gómez, G. (2014). Modalidades participativas de evaluación. Un análisis de la percepción del profesorado y de los estudiantes universitarios. *Revista de Investigación Educativa*, 32(2), 339-361. <https://doi.org/10.1080/02602930903221444>
- López Lozano, L. y Solís Ramírez, E. (2016). Con qué evalúan los estudiantes de magisterio en formación. *Campo Abierto. Revista de Educación*, 35(1), 55-67. <https://dialnet.unirioja.es/servlet/articulo?codigo=5787079>
- López Pastor, V. (2005). La participación del alumnado en la evaluación. La autoevaluación, la coevaluación y la evaluación compartida. *Tándem*, 17, 1-8.

- Lukas, J. F., Santiago, K., Lizasoain, L. y Etxebarria, J. (2017). Percepciones del alumnado universitario sobre su evaluación. *Bordón*, 69(1), 103-122. <https://doi.org/10.13042/Bordon.2016.43843>
- Mingorance, C. y Calvo, A. (2013). Los resultados de los estudiantes en un proceso de evaluación con metodologías distintas. *Revista de Investigación Educativa*, 31(1), 275-293. <https://doi.org/10.6018/rie.31.1.153291>
- Moro, A. I. (2016): La evaluación de las técnicas de evaluación en la enseñanza universitaria: la experiencias de macroeconomía. *Cultura y Educación*, 28(4), 843-862.
- Palacios Picos, A. y López-Pastor, V.M. (2013). Haz lo que yo digo pero no lo que yo hago: sistemas de evaluación del alumnado en la formación inicial del profesorado. *Revista de Educación*, 361, 279-305. <http://doi.org/10.4438/1988-592X-RE-2011-361-143>.
- Pascual Gómez, I., Lorenzo Llamas, E. M^a. y Monge López, C. (2015). Análisis de validez en la evaluación entre iguales: un estudio en educación superior. *RELIEVE*, 21(1), 1-17. <https://doi.org/10.7203/relieve.21.1.5168>
- Pérez Juste, R. (2016). Quo vadis, evaluación? Reflexiones pedagógicas en torno a un tema tan manido como relevante. *Revista de Investigación Educativa*, 34(1), 13-30. <https://doi.org/10.6018/rie.34.1.239381>
- Quesada, V., Rodríguez Gómez, G. e Ibarra Sáiz, M.S. (2017). Planificación e innovación de la evaluación en educación superior: la perspectiva del profesorado. *Revista de Investigación Educativa*, 35(1), 53-70. <https://doi.org/10.6018/rie.35.1.239261>
- Rodríguez Espinosa, H., Retrespo, L. F. y Luna, G. C. (2016). Percepción del estudiantado sobre la evaluación del aprendizaje en la educación superior. *Revista Electrónica Educare*, 20(3), 1-17. <https://doi.org/10.15359/ree.20-3.18>
- Rodríguez Gómez, G., Ibarra Sáiz, M^a. S. y García Jiménez, E. (2013). Autoevaluación, evaluación entre iguales y coevaluación: conceptualización y práctica en las universidades españolas. *Revista de Investigación en Educación*, 11(2), 198-210. <http://reined.webs.uvigo.es/ojs/index.php/reined/article/view/708>
- Rodríguez Fuentes, A., Gallego Ortega y Figueroa, S. (2018). Heteroevaluación, coevaluación y autoevaluación universitaria. *Revista Iberoamericana de Educación Superior*, en prensa.
- Saiz, M^a. C. y Bol, A. (2014). Aprendizaje basado en la evaluación mediante rúbricas en educación superior. *Suma Psicológica*, 21(1), 28-35. [https://doi.org/10.1016/S0121-4381\(14\)70004-9](https://doi.org/10.1016/S0121-4381(14)70004-9)
- Saiz, M^a. C. y Payo, R.J. (2012). Auto-percepción del conocimiento en educación superior. *Revista iberoamericana de Psicología y Salud*, 3(2), 159-174.
- Sluijsmans, D., Dochy, F. y Moerkerke, G. (1999). Creating a Learning Environment by Using Self-, Peer - and Co-Assessment. *Learning Environment Research*, 1, 293-319. <https://doi.org/10.1023/A:1009932704458>
- Stake, R. (2010). *Investigación con estudio de casos*. Madrid: Morata.

Wheater, C. P., Langan, A. M. y Dunleavy, P. J. (2005). Students assessing students: case studies of peer assessment". *Planet*, 15, 1-15. <https://doi.org/10.11120/plan.2005.00150013>

Zaragoza, J., Luis, J.C. y Manrique, J.C. (2009). Experiencia de innovación en docencia universitaria: resultado de la aplicación de sistemas de evaluación formativa. *REDU. Revista de Docencia Universitaria*, 7(4), 1-33. <https://doi.org/10.4995/redu.2009.6232>

Artículo concluido el 5 de Septiembre de 2017

Gallego Ortega, J. L., Rodríguez Fuentes, A. (2017). Confiabilidad de las valoraciones de los estudiantes. Estudio de sus competencias evaluativas. *REDU. Revista de Docencia Universitaria*, 15(2), 349-366.

<https://doi.org/10.4995/redu.2017.8780>

José Luis Gallego Ortega

Universidad de Granada
Didáctica y Organización Escolar
jlgalleg@ugr.es

Licenciado en Filología Hispánica y Doctor en Ciencias de la Educación, ha ejercido de maestro en todos los niveles del sistema educativo y actualmente como Profesor Titular de la Facultad de Ciencias de la Educación de la Universidad de Granada. Dirige el Grupo de investigación I.C.E. (Investigación en Comunicación Educativa, HUM-871) de la Universidad de Granada. Sus principales líneas de investigación son la Educación Especial y las Dificultades de Lenguaje oral y escrito. Sobre estos temas ha publicado una treintena de libros, divulgados por editoriales de prestigio internacional (como Aljibe, Pirámide, EOS, McGraw Hill, Vicens Vices, etc.); y cerca de un centenar de artículos, editados en distintas revistas internacionales (americanas y europeas) y nacionales, así como numerosos capítulos de libros publicados en otros tantos textos académicos, con una participación igualmente relevante y destacable en Congresos, Jornadas y eventos académicos y formativos de ámbito nacional e internacional.

Antonio Rodríguez Fuentes

Universidad de Granada
Didáctica y Organización Escolar
arfuentes@ugr.es

Licenciado en Pedagogía, Dr. en Ciencias de la Educación y Profesor Titular de la Universidad de Granada, especializado en la atención a la diversidad en la Escuela Inclusiva. Autor de libros en editoriales especializadas, internacionales y de prestigio, como Arial, Aljibe, EOS, Octaedro, Pirámide, principalmente; así como de artículos en

diversas revistas igualmente prestigiosas e internacionales (europeas y americanas). Ponente y profesor invitado en universidades extranjeras de Colombia, Francia, Italia, Marruecos, Panamá y Portugal, además de españolas. Entre sus líneas de investigación resaltan la atención al alumnado con discapacidades en el marco de una Educación Inclusiva y una Escuela para Todos, principalmente en la adquisición de habilidades instrumentales básicas y académicas básicas. Forma parte de diferentes revistas del campo de la educación, tanto de su panorama nacional (como Enseñanza & Teaching y Revista de la Universidad de Granada), como de otros contextos iberoamericanos (como Propósitos y Representaciones, RETOS XXI) y es revisor de otras muchas, también tanto nacionales (Profesorado, Revista de Educación, Revista de Docencia Universitaria, Etic@net, Tendencias Pedagógicas, Revista de Investigación Educativa, Ámbitos. Revista Internacional de Comunicación, Revista InnoEduca) como de otros contextos foráneos (México: Revista Mexicana de Investigación Educativa; Colombia: Revista Psicología del Caribe, Revista Oniriquia y Revista Educación y Educadores; Venezuela: Revista Multiciencias de la Universidad del Zulia; Northem Ireland: International Journal of Educational Research; Ecuador: Revista Estoa de Arquitectura y Urbanismo). Ejerce como revisor de textos de la Editorial EOS de textos de psicología y educación.