

# INFLUENCIA DE UN PROCESO DE AUTOEVALUACIÓN, COEVALUACIÓN Y EVALUACIÓN EN LA FORMACIÓN DE MAESTROS DE PRIMARIA

INFLUENCE OF A SELF-ASSESSMENT, PEER-ASSESSMENT  
AND ASSESSMENT PROCESS IN PRIMARY SCHOOL  
TEACHERS TRAINING

CÁCERES, M. J., CHAMOSO, J. M.  
*Universidad de Salamanca*

## RESUMEN

La asignatura Matemáticas y su Didáctica III del Grado en Maestro de Educación Primaria en la Universidad de Salamanca incluye el diseño y desarrollo de un Proyecto Estadístico, así como su posible aplicación didáctica, que los estudiantes realizan de forma grupal y autónoma contando con la formación, seguimiento y tutoría del profesor. En una fase intermedia cada grupo entrega un Avance del Proyecto Estadístico y se somete a un proceso de autoevaluación, coevaluación y evaluación del profesor, siguiendo la rúbrica de valoración presentada al inicio de la asignatura, para desarrollar la competencia evaluativa como aspecto integrante del conocimiento didáctico del contenido. Los resultados de las valoraciones fueron consistentes en la mayoría de los aspectos considerados y las discrepancias entre las valoraciones de los estudiantes y del profesor facilitaron la comprensión de los apartados del Proyecto Estadístico y su correspondiente mejora.

Palabras clave: *formación de docentes, evaluación para el aprendizaje, autoevaluación, coevaluación, proyecto estadístico.*

Cáceres, M.J. y Chamoso, J.M. (2019). Influencia de un proceso de autoevaluación, coevaluación y evaluación en la formación de profesores de primaria. En E. Badillo, N. Climent, C. Fernández y M. T. González (Eds.), *Investigación sobre el profesor de matemáticas: formación, práctica de aula, conocimiento y competencia profesional* (pp. 351-372). Salamanca: Ediciones Universidad Salamanca.

## ABSTRACT

The subject Mathematics and its Didactics III of Primary Education Teacher degree at the University of Salamanca includes the design and development of a Statistical Project, as well as its possible didactic application, which students self-perform in groups, with the teacher's training, follow-up and tutoring. In an intermediate phase each group delivers a Statistical Project Progress report subject to a self-evaluation, peer-evaluation and teacher's evaluation, following the assessment criteria presented at the beginning of the subject to develop the evaluative competence as an integral aspect of the content knowledge for teaching. The assessment results were consistent in most aspects and the discrepancies between the students' and the teacher's assessments made easier the understanding of the sections of the Statistical Project and their corresponding improvement.

*Keywords: teacher training, assessment for learning, self-assessment, peer-assessment, statistical project.*

## INTRODUCCIÓN

ESTE TRABAJO se desarrolla en un contexto de formación de maestros de Primaria en Matemáticas durante los estudios de Grado. Se pretende que desarrollen competencias tanto de conocimiento matemático como de conocimiento didáctico del contenido (Ball, Thames y Phelps, 2008); es decir, se considera que deben adquirir un conocimiento teórico (saber), así como destrezas suficientes para impartir los contenidos (saber hacer), lo que puede organizarse en competencias matemáticas y profesionales (ver Cáceres, Chamoso y Azcárate, 2010).

Un aspecto importante del conocimiento didáctico del contenido es la evaluación. Los tipos de evaluación que se lleven a cabo en el aula pueden desempeñar un papel fundamental en el aprendizaje de los estudiantes, a la vez que proporcionar evidencias del progreso en ese aprendizaje (Boud, Cohen y Sampson, 1999). Para que un proceso involucrado en tareas formativas cobre importancia es adecuado que se considere una evaluación formativa que fomente la reflexión y permita la revisión del propio trabajo (Williams, 2017; Cáceres et al., 2010; Chamoso y Cáceres, 2009). Para ello sería conveniente involucrar a los estudiantes para maestro en el proceso evaluativo y pasar de la evaluación de aprendizajes a la evaluación para el aprendizaje. Una posibilidad es promover la evaluación entre iguales con autoevaluación y coevaluación unido a la evaluación del profesor (Miedijensky y Tal, 2016). La evolución de los trabajos valorados dará cuenta de la parte formativa de este tipo de evaluación.

Este proceso se puede desarrollar de diversas formas. En este trabajo se intenta que se desarrolle cuando los futuros docentes están elaborando un Proyecto Estadístico, objetivo general de una materia que deben cursar durante su formación inicial. En una sociedad de la información como la actual, la vida cotidiana de las personas no escapa a la presencia elevada de datos cuantitativos y estadísticas que se deben saber gestionar. Ello supone poseer un sólido conocimiento que involucre

habilidades para el tratamiento de datos, que incluye resumir, analizar, interpretar y divulgar la información (Watson, 2002). Por ello, la formación actual de docentes requiere su capacitación para diseñar y evaluar tareas centradas en el análisis e interpretación de datos reales, así como conocer y experimentar procesos de evaluación de los aprendizajes.

## FUNDAMENTACIÓN TEÓRICA

### EL DESARROLLO DE COMPETENCIAS EN LA FORMACIÓN DE DOCENTES

De forma general se recomienda que, en la enseñanza universitaria, se desarrollen competencias generales o transversales como son la resolución de problemas, la formulación de preguntas, la búsqueda y uso de información relevante, la realización de juicios reflexivos y el desarrollo de pensamiento crítico y aprendizaje autónomo (Ibarra, Rodríguez y Gómez, 2012). En el ámbito de la formación del profesorado, si bien existe abundante investigación sobre el conocimiento que debe poseer un docente de Matemáticas, esta no está habitualmente centrada en el desarrollo de la competencia evaluativa del conocimiento estadístico.

En cuanto al desarrollo del conocimiento estadístico, Garfield y Everson (2009) recomendaron enfatizar el alfabetismo y pensamiento estadístico, utilizar datos reales, estimular la comprensión conceptual, promover el aprendizaje activo en el aula, utilizar tecnología, y evaluar para mejorar y valorar el aprendizaje del estudiante. Para ello, algunos autores aconsejaron la utilización de Proyectos Estadísticos durante la formación universitaria de los futuros maestros (Batanero y Díaz, 2011; Browning, Goss y Smith, 2014; Franklin et al., 2005).

El desarrollo de un Proyecto Estadístico comienza con el proceso de búsqueda de información para satisfacer la curiosidad y los intereses de los estudiantes y, en función de dichos intereses, formular preguntas sobre lo que se desea investigar y para cuya respuesta se precisa recoger datos, analizarlos e interpretarlos (Batanero y Díaz, 2011; Burrill y Camden, 2006; Dierker et al., 2016; Franklin, et al. 2005). Hacer que los estudiantes recopilen y analicen datos creará una red de significados que promoverá la construcción del conocimiento estadístico y su aprendizaje en una investigación creativa a partir de su planificación, descripción del fenómeno, elección del tipo de muestreo y selección del método estadístico de análisis (da Silva, Porciúncula y Pinto, 2014).

Esos ambientes de aprendizaje auténticos, con metodologías flexibles, abiertas a la colaboración y participación de los estudiantes a partir de sus propias inquietudes, recomiendan que la evaluación sea realizada entre iguales y unida a la del profesor (Bretones, 2008; Keppell, Au, Ma, y Chan, 2006). Por ello es conveniente el conocimiento por parte de los estudiantes para maestro de este tipo de evaluación.

## EVALUACIÓN, AUTOEVALUACIÓN Y COEVALUACIÓN

La evaluación entre iguales, entendida como autoevaluación y coevaluación, tiene un largo recorrido en diversos países y materias en la enseñanza universitaria y, en concreto, en la formación de maestros (Cáceres et al., 2010; Chamoso y Cáceres, 2009; Ibarra et al., 2012). Por otra parte, si bien es cierto que en la práctica de las aulas universitarias sigue siendo frecuente que el peso de la evaluación recaiga sobre el profesor, progresivamente se intenta conceder mayor protagonismo a los estudiantes, a la retroalimentación obtenida mediante procesos de evaluación formativa, entendida como información útil que permita una mejora de lo realizado o aprendido (Carless, Joughin y Mok, 2006) y al desarrollo de tareas auténticas mediante la utilización de estrategias de evaluación que favorezcan oportunidades de aprendizaje de los estudiantes (Keppell et al., 2006). La retroalimentación entre pares puede ser más influyente que la retroalimentación del docente y permite entender que la integración de la autoevaluación y coevaluación en el proceso de evaluación puede favorecer la participación, la motivación y la incentivación de los estudiantes (Chang, Tseng y Lou, 2012).

Por otra parte, la autoevaluación es particularmente beneficiosa para los estudiantes, ya que no solo los mantiene involucrados, interesados y motivados en el proceso, sino que, además, fomenta la autorreflexión, la responsabilidad y la valoración del propio trabajo con la toma de conciencia de fortalezas y debilidades, lo que ayuda a mejorar el conocimiento del contenido (Cáceres, Chamoso y Cárdenas, 2015; Chang et al., 2012; Miller, 2003; Sadler y Good, 2006). La coevaluación fomenta la autorreflexión, permite que los estudiantes vean los trabajos de otros, reciban ideas y sugerencias de compañeros, tomen conciencia de la calidad de su trabajo comparando con los de otros y adapten su trabajo de acuerdo con los comentarios que reciban de sus compañeros (Bouzidi y Jaillet, 2009; Chang et al., 2012; Chen, 2010).

Algunos investigadores han cuestionado la validez y la fiabilidad de la autoevaluación y coevaluación en el sentido de que sus resultados reflejen realmente los logros de aprendizaje (Chang et al., 2012; Ibarra et al., 2012). Así, por ejemplo, en el caso de la autoevaluación, parece que los estudiantes más débiles tienden a infravalorarse mientras que los estudiantes más fuertes tienden a sobrevalorarse (Lejik y Wyvill, 2001; Sadler y Good, 2006; Wilmot y Crawford, 2005). Investigaciones como la de Lin, Liu y Yuan (2001a, b) mostraron que la coevaluación tiene mejores niveles de validez y confiabilidad que la autoevaluación, aunque también puede estar afectada por factores como sobrevaloración por relaciones de amistad, liderazgo en el grupo y personas que se benefician del trabajo de otros (Ibarra et al., 2012).

Un instrumento que facilita labores evaluativas, como pueden ser la comprensión y el uso de los comentarios de retroalimentación y la autorregulación del

aprendizaje de los estudiantes, son las rúbricas de evaluación (Cáceres y Chamoso, 2015; Jönsson y Panadero, 2017). Para realizar evaluación formativa se recomiendan rúbricas analíticas, que permiten una evaluación pormenorizada de los diversos aspectos clave del proceso (Nitko, 2001). En algunos trabajos se aprecia la posibilidad de que, en la valoración con rúbricas de evaluación, los docentes utilicen criterios estrictos de puntuación mientras que los estudiantes tengan criterios más laxos, de manera que la valoración de los estudiantes es más alta que la del docente y, en unos casos, la autoevaluación valora más alto que la coevaluación (Lin et al., 2001a, b), aunque, en otros, ocurre lo contrario (por ejemplo, Sadler y Good, 2006).

La investigación muestra gran variedad de matices en relación con los resultados de autoevaluación, coevaluación y la evaluación del docente. En algunos casos se muestra la consistencia entre las tres formas de evaluación (Sadler y Good, 2006; Sung, Chang, Chiou y Hou, 2005), pero en otros se obtiene consistencia entre autoevaluación y evaluación del docente (Bouzidi y Jailliet, 2009; Cho, Schunn y Wilson, 2006; Sadler y Good, 2006; Sung et al., 2005; Tsai y Liang, 2009; Tseng y Tsai, 2007), al contrario que en el trabajo de Knowles, Holton III y Swanson (2005). Respecto a la autoevaluación y coevaluación se han detectado inconsistencias entre coevaluación y evaluación del docente debidas a los diversos niveles de conocimiento de los estudiantes, la utilización de rúbricas de evaluación o la utilización de diferentes procedimientos de evaluación (Chang et al., 2012; Chen, 2010).

La mayor parte de estos trabajos se realizaron a partir de la valoración de portafolios de aprendizaje, en muchos casos electrónicos, pero son escasos los que analizan la influencia de una sesión formativa sobre el uso de rúbricas de evaluación, mediante un proceso de evaluación entre iguales, donde se compartan públicamente las valoraciones de las producciones de los alumnos que permita la reconsideración del propio trabajo durante la formación de docentes en el sentido de Ibarra et al. (2012). En este trabajo se avanza en ese sentido, en concreto, en un contexto de formación de maestros de Primaria que, en grupos, elaboran un Proyecto Estadístico. En un momento intermedio de su desarrollo, los estudiantes presentaron el Avance del Proyecto Estadístico que, a partir de una rúbrica de valoración, fue valorado por ellos mismos (*Autoevaluación*), sus compañeros (*Coevaluación*) y el profesor (*Evaluación*). El objetivo de este trabajo es analizar:

- las consistencias y discrepancias entre las valoraciones dadas a cada Avance del Proyecto Estadístico mediante la *Autoevaluación*, la *Coevaluación* y la *Evaluación*
- el tratamiento del Avance del Proyecto Estadístico de cada grupo a partir del propio trabajo y de las valoraciones realizadas al mismo.

## METODOLOGÍA

### CONTEXTO Y MUESTRA

La muestra quedó conformada por 46 grupos (17 grupos en el curso 2015-16, 12 en 2016-17 y 17 en 2017-18) de 3-5 estudiantes que cursaban la asignatura *Matemáticas y su Didáctica III* (6 ECTS), integrada en el plan de estudios del Grado en Maestro de Educación Primaria.

Los contenidos de la asignatura fueron los relacionados con la enseñanza de Estadística y Probabilidad en Primaria. Para mostrar un adecuado desarrollo de las competencias establecidas en el plan docente, los estudiantes, en grupos, debían ser capaces de identificar las etapas del desarrollo de un Proyecto Estadístico (PE) y realizar uno, así como diseñar y presentar una secuencia de aula relacionada con su PE de una forma similar a como lo harían en el desempeño de su futura labor profesional. Además, cada estudiante, de forma individual, debía realizar otras tareas complementarias a lo largo del curso enfocadas a facilitar tanto el desarrollo del PE como su implementación. Por otra parte, como herramienta de apoyo al proyecto se recurrió a la plataforma virtual *Studium* de la Universidad de Salamanca (más detalle, González y Chamoso, 2015).

La formación se desarrolló a lo largo de 15 semanas, en 4 horas semanales distribuidas en 2 sesiones de 2 horas de duración, organizadas en: sesiones formativas (una semanal), con todos los estudiantes matriculados, en las que se trabajaron tanto los contenidos de Estadística y Probabilidad para su enseñanza en Primaria como los relacionados con el desarrollo de un PE; y sesiones prácticas (una semanal), donde cada grupo de estudiantes avanzaba en su propio PE en dos aulas diferentes, en cada una de las cuales se encontraba la mitad de los estudiantes matriculados, con un profesor que atendía sus avances y dudas ya fuera de manera puntual o de forma más profunda.

La elaboración del PE por cada grupo de estudiantes consistía en seleccionar un objetivo de interés y, relacionado con él, buscar documentación sobre el mismo que permitiera formular preguntas que pudieran ser respondidas mediante la utilización de herramientas estadísticas, para lo que era necesario recoger y organizar datos, analizar los datos obtenidos e interpretar los resultados. Al final del curso, en las sesiones de clase usuales, cada grupo debía entregar y presentar su PE ante sus compañeros y profesores.

Para valorar los PE se utilizó una rúbrica analítica que los estudiantes conocían desde el comienzo del curso. Dicha rúbrica se creó a partir de la literatura especializada (Batanero y Díaz, 2011; Franklin et al. 2005; Garfield y Everson, 2009) y se validó previamente mediante su aplicación a los PE de 15 grupos de estudiantes diferentes de los considerados en este trabajo, a través de un análisis

de concordancia o acuerdo intercodificadores con la participación independiente de dos profesores (Kappa de Cohen, 0,93; las discrepancias ayudaron a mejorar la redacción de algunos indicadores). En esencia, la rúbrica consideraba 5 aspectos directamente relacionados con la elaboración de un PE, cada uno de los cuales se valoraba de 1 a 4 en función de la calidad alcanzada: Planteamiento y justificación del problema y objetivos (*Problema*), Búsqueda de información (*Información*), Recogida y organización de datos (*Datos*), Análisis de datos (*Análisis*) e Interpretación de los datos (*Interpretación*) (Tabla 1):

Tabla 1. Rúbrica de valoración de los APE y MAPE

	4	3	2	1
Planteamiento y justificación del problema y objetivos ( <i>Problema</i> )	El tema es pertinente, el problema se justifica adecuadamente y se formulan objetivos medibles estadísticamente	El tema es pertinente, el problema se justifica vagamente o se formulan objetivos que no son medibles estadísticamente	El tema es pertinente pero no se justifica adecuadamente su elección o se formulan objetivos para los que no es necesario aplicar la estadística	El tema no es pertinente y no se justifica su elección. No se formulan objetivos o se formulan objetivos ambiguos
Búsqueda de información ( <i>Información</i> )	Se utilizan suficientes fuentes de información adecuadas al tema de estudio y se incluyen en la bibliografía	Se utilizan algunas fuentes de información adecuadas al tema de estudio y se incluyen en la bibliografía	Se utilizan algunas fuentes de información adecuadas al tema de estudio y no se incluyen en la bibliografía	No se muestran evidencias de haber utilizado fuentes de información.
Recogida y organización de datos ( <i>Datos</i> )	Se indica cómo se recogen los datos, los instrumentos de recogida y la explicación clara de su organización	Se indica cómo se recogen los datos y los instrumentos de recogida, pero no cómo se organizan	Únicamente se indica cómo se han recogido los datos	No se indica el proceso de recolección de datos
Análisis de Datos ( <i>Análisis</i> )	Las medidas utilizadas, las gráficas o tablas son las adecuadas para los objetivos planteados y se elaboran correctamente	Las medidas utilizadas, las gráficas y las tablas son las adecuadas, pero se cometen errores en su elaboración	Las medidas, las tablas o las gráficas no son adecuadas	No se precisan las medidas o las tablas y gráficas no tienen sentido para los objetivos planteados

	4	3	2	1
Interpretación de los datos ( <i>Interpretación</i> )	La interpretación es correcta y completa para los objetivos planteados. Se realizan inferencias a partir de los datos	La interpretación es correcta y adecuada para los objetivos planteados, pero no se realizan inferencias a partir de los datos	La interpretación o es correcta o no es adecuada para los objetivos planteados y no se realizan inferencias a partir de los datos	La interpretación no guarda relación con los objetivos planteados

Para facilitar el desarrollo del PE, en la quinta semana de curso y en las condiciones acordadas al inicio del mismo, cada grupo de estudiantes debía entregar los Avances de su Proyecto Estadístico (APE), los cuales debía presentar públicamente ante sus compañeros y los profesores para que aportaran sugerencias o recomendaciones de mejora en cada caso. Los APE de cada grupo fueron valorados por el profesor (*Evaluación*) y en una sesión posterior se realizó una intervención educativa para facilitar la reflexión de cada grupo de estudiantes sobre su propio APE que permitiera avanzar en su desarrollo. Para ello, cada grupo participó de la siguiente forma:

a) Familiarización con la rúbrica de valoración:

1. El grupo aplicó la rúbrica a su propio APE (20 minutos; *Autoevaluación*; entrega por escrito en Studium).
2. El grupo aplicó la rúbrica a 3 APE desarrollados por compañeros de cursos previos (20 minutos). Los resultados se comentaron en el aula con el objetivo de que sugirieran posibilidades para que cada grupo pudiera mejorar su propio APE (20 minutos).

b) Aplicación de la rúbrica:

3. El grupo aplicó la rúbrica a 5 APE desarrollados por sus compañeros (20 minutos; *Coevaluación*; entrega por escrito en Studium).
4. El grupo validó o modificó justificadamente la valoración de su propio APE, incluyendo la calificación (20 minutos; entrega por escrito en Studium).

Posteriormente, como parte de su trabajo de la asignatura, cada grupo de estudiantes reconsideró su propio APE e introdujo las mejoras que consideró oportunas, entregándolo de nuevo (Mejora de los Avances del Proyecto Estadístico, MAPE).

Los APE y MAPE de los estudiantes no consideraron todos los aspectos establecidos en la elaboración de los PE, por lo que se valoraron únicamente los tres primeros de la rúbrica (*Problema, Información y Datos*).

La influencia de esta intervención educativa en la evolución del trabajo de los estudiantes para maestro se estudió en dos sentidos. En primer lugar, se analizaron



la consistencia y discrepancias entre las tres formas de valoración en los APE de cada grupo (*Autoevaluación*, *Coevaluación* y *Evaluación*) y, en segundo lugar, se analizó el progreso en cada MAPE respecto a su correspondiente APE. En este trabajo entendemos que dos formas de valoración son consistentes cuando cada resultado numérico asignado a partir de la rúbrica de valoración (en adelante valoración) mediante una forma de valoración se mantiene con otra y entendemos las discrepancias entre las formas de valoración como las diferencias entre valoraciones de un mismo trabajo (Chang et al., 2012).

Los datos fueron las valoraciones obtenidas por los distintos grupos en sus APE mediante *Autoevaluación* (valoración asignada por cada grupo a su propio APE), *Coevaluación* (media aritmética de las valoraciones asignadas por los grupos que valoraron el mismo APE) y *Evaluación* (valoración asignada por el profesor a cada APE) y las valoraciones obtenidas por cada grupo en MAPE mediante *Evaluación*.

Se realizaron estudios estadísticos con pruebas no paramétricas de correlaciones y comparación de valoraciones y descriptivos de las diferencias entre valoraciones. En concreto, en primer lugar, se estudió la consistencia entre las tres formas de valoración a partir de las correlaciones dos a dos de las valoraciones obtenidas en los APE en *Autoevaluación*, *Coevaluación* y *Evaluación* mediante el coeficiente de Spearman (puesto que la muestra no sigue las condiciones de normalidad); además, se realizó el estudio de las discrepancias entre las tres formas de valoración mediante la comparación dos a dos de las valoraciones obtenidas por los grupos en *Autoevaluación*, *Coevaluación* y *Evaluación* con la prueba de rangos con signo de Wilcoxon para muestras relacionadas (prueba no paramétrica equiparable a la prueba t para muestras dependientes), tanto de forma Global (media aritmética de los aspectos considerados) como para cada aspecto *Problema*, *Información* y *Datos*, y se estudiaron las diferencias entre las valoraciones otorgadas a los APE con cada forma de valoración. En segundo lugar, para analizar la consistencia de las valoraciones obtenidas por cada grupo mediante la *Evaluación* en los MAPE y APE, tanto globalmente como para cada aspecto considerado en la rúbrica de valoración, se aplicó la prueba de rangos con signo de Wilcoxon y el estudio descriptivo de las diferencias de valoración de cada MAPE respecto a su correspondiente APE.

## RESULTADOS

A continuación, se presentan los resultados obtenidos para cada uno de los objetivos presentados. En primer lugar, se muestran las consistencias y discrepancias dos a dos entre las valoraciones obtenidas por los grupos a partir de la rúbrica de valoración mediante *Autoevaluación*, *Coevaluación* y *Evaluación* y, en segundo lugar, se considera el tratamiento que cada grupo realizó en su MAPE fruto de la reflexión y de las valoraciones realizadas en su respectivo APE.

CONSISTENCIAS Y DISCREPANCIAS ENTRE VALORACIONES SEGÚN LA FORMA DE VALORACIÓN

*Comparación de resultados en Autoevaluación y Coevaluación*

La correlación entre la valoración Global obtenida por los grupos en *Autoevaluación* y *Coevaluación* en sus APE es positiva y significativa (coeficiente de correlación de Spearman), al igual que entre las valoraciones de *Problema* e *Información*, lo que sugiere un alto grado de consistencia entre las valoraciones en APE mediante *Autoevaluación* y *Coevaluación*; no se dieron diferencias significativas para la prueba de Wilcoxon en los resultados obtenidos en la mayoría de aspectos, lo que indica un bajo grado de discrepancias (Tabla 2). Esto hace pensar que los estudiantes entendieron de la misma forma los diversos indicadores de la rúbrica.

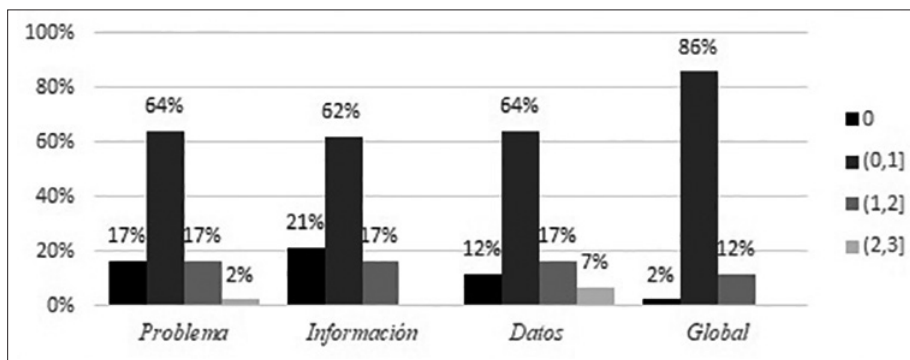
Tabla 2. Correlación de Spearman y prueba de Wilcoxon para *Autoevaluación* y *Coevaluación*

	<i>Coefficiente de Correlación (sig.)</i>	<i>p de Wilcoxon</i>
<i>Problema</i>	0,373 (0,015*)	0,041*
<i>Información</i>	0,554 (0,000**)	0,176
<i>Datos</i>	0,282 (0,071)	0,668
Global	0,408 (0,007*)	0,472

\* $p < 0,05$  \*\* $p < 0,01$

En la valoración Global, el 64% de los grupos obtuvo, en *Autoevaluación*, valoraciones superiores o iguales que en *Coevaluación* y, en cada aspecto, las valoraciones en *Autoevaluación* fueron superiores o iguales que en *Coevaluación* para el 79% de los grupos en *Problema* y el 69% en *Información*; sin embargo, para el 55% de los grupos fueron inferiores en *Datos*. La Figura 1 muestra que en este aspecto es donde se dio el menor porcentaje de coincidencias entre valoraciones (12%) y el mayor porcentaje entre diferencias superiores a 2 puntos (12%). Parece que los estudiantes tuvieron dificultades para valorar sus propios trabajos y consideraron más su esfuerzo que lo que realmente supieron transmitir en el documento escrito, de forma que premiaron o penalizaron aspectos que sus compañeros no fueron capaces de detectar.

Figura 1. Porcentajes de diferencias, en valor absoluto, para *Autoevaluación* y *Coevaluación*



#### *Comparación de resultados en Autoevaluación y Evaluación*

La correlación entre la valoración Global obtenida por los grupos en *Autoevaluación* y *Evaluación*, en cada respectivo APE, es positiva y significativa (coeficiente de correlación de Spearman), lo que sugiere un alto grado de consistencia global, aunque, atendiendo a cada aspecto, la correlación solo fue significativa en *Datos*; además, entre las valoraciones en APE mediante *Autoevaluación* y *Evaluación* se dieron diferencias significativas para la prueba de Wilcoxon tanto en la valoración Global como para cada aspecto, lo que refleja grandes discrepancias entre las valoraciones asignadas por cada grupo a su propio APE y las del profesor (Tabla 3). Esto hace pensar que el grado de consistencia global es engañoso debido, quizás, a que la media de valoraciones obtenidas en los diversos aspectos compensa desigualdades en las valoraciones a nivel individual de cada uno de ellos, y a que los estudiantes no entendían la rúbrica de la misma forma que el profesor.

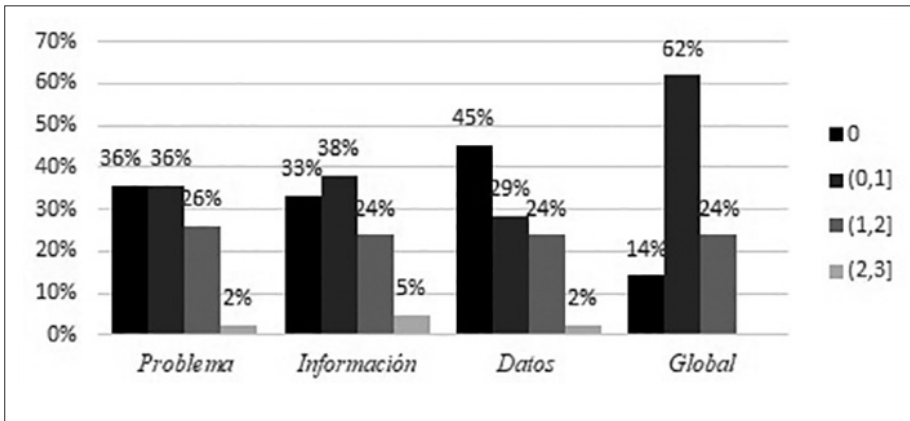
Tabla 3. Correlación de Spearman y prueba de Wilcoxon para *Evaluación* y *Autoevaluación*

	<i>Coefficiente de Correlación (sig.)</i>	<i>p de Wilcoxon</i>
<i>Problema</i>	0,254 (0,104)	0,000**
<i>Información</i>	0,277 (0,076)	0,000**
<i>Datos</i>	0,353 (0,022*)	0,008**
Global	0,408 (0,007**)	0,000**

\* $p < 0,05$  \*\* $p < 0,01$

En la valoración Global, el 95% de los grupos obtuvo, en *Autoevaluación*, valoraciones superiores o iguales que en *Evaluación* y, en cada aspecto, las valoraciones en *Autoevaluación* fueron superiores o iguales que en *Evaluación* para el 93% de los grupos en *Problema*, el 95% en *Información* y el 88% en *Datos*. En La Figura 2 llama la atención el elevado porcentaje de coincidencias en los tres aspectos, pero también el alto porcentaje de grupos que obtuvo valoraciones en su APE con diferencias entre 1 y 3 puntos, lo que lleva a grandes discrepancias entre las valoraciones. De nuevo, parece que los estudiantes valoraron aspectos, quizás trabajados, pero que no habían sabido plasmar en los documentos valorados.

Figura 2. Porcentajes de diferencias, en valor absoluto, para *Autoevaluación* y *Evaluación*



#### *Comparación de resultados en Coevaluación y Evaluación*

La correlación entre la valoración Global obtenida por los grupos en *Coevaluación* y *Evaluación* en sus APE es positiva y significativa (coeficiente de correlación de Spearman) al igual que en *Información* y *Datos*, lo que sugiere un alto grado de consistencia entre las valoraciones en APE mediante *Coevaluación* y *Evaluación*; las diferencias significativas para la prueba de Wilcoxon tanto en la valoración Global como en cada aspecto refleja grandes discrepancias entre las valoraciones asignadas por los estudiantes al APE de sus compañeros y las del profesor (Tabla 4). En este caso, la consistencia en las valoraciones lleva a pensar que distinguieron los niveles establecidos en la rúbrica de evaluación, pero que los valoraron más alto que el profesor, lo que lleva a dichas discrepancias.

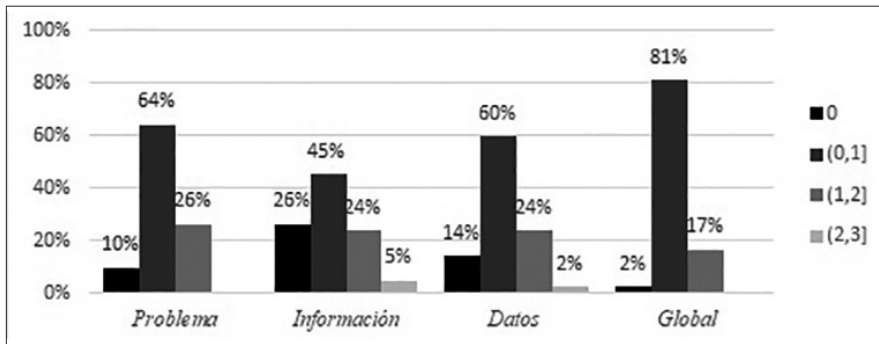
Tabla 4. Correlación de Spearman y prueba de Wilcoxon para *Evaluación* y *Coevaluación*

	<i>Coefficiente de Correlación (sig.)</i>	<i>p de Wilcoxon</i>
<i>Problema</i>	0,144 (0,362)	0,002**
<i>Información</i>	0,402 (0,008**)	0,000**
<i>Datos</i>	0,384 (0,022*)	0,000**
Global	0,510 (0,001**)	0,000**

\*p<0,05 \*\*p<0,01

De hecho, en la valoración Global, el 93% de los grupos obtuvo, en *Coevaluación*, valoraciones superiores o iguales que en *Evaluación* y, en cada aspecto, las valoraciones en *Coevaluación* fueron superiores o iguales que en *Evaluación* para el 76% de los grupos en *Problema*, el 93% en *Información* y el 81% en *Datos*. La Figura 3 muestra el alto porcentaje de grupos que obtuvo valoraciones en su APE con diferencias entre 1 y 3 puntos para *Coevaluación* y *Evaluación*.

Figura 3. Porcentajes de diferencias, en valor absoluto, para *Coevaluación* y *Evaluación*



En resumen, los resultados fueron consistentes entre las tres formas de valoración, aunque en menor medida entre *Autoevaluación* y *Evaluación*, lo que quiere decir que, aunque no coincidan en valor, generalmente los grupos que obtuvieron las valoraciones más altas o más bajas en una forma de valoración fueron, respectivamente, los mismos que lo hicieron en las otras dos. Las principales discrepancias se dieron entre las valoraciones realizadas por los estudiantes y las realizadas por el profesor, especialmente en *Coevaluación*, donde las coincidencias con las valoraciones del profesor fueron siempre inferiores al 15%, salvo en *Información* donde fueron el 26%.

## TRATAMIENTO EN MAPE DE LAS VALORACIONES REALIZADAS EN APE

La correlación entre la valoración Global obtenida por los grupos mediante *Evaluación* en APE y MAPE es positiva y significativa (coeficiente de correlación de Spearman), al igual que en *Problema* e *Información*, lo que sugiere un alto grado de consistencia entre las valoraciones obtenidas por cada grupo en APE y MAPE mediante *Evaluación*. Esto parece indicar que los grupos, a partir del proceso de evaluación formativa desarrollado con sus APE, modificaron sustancialmente todos los aspectos, manteniéndose que las mayores y menores valoraciones en MAPE tendieron a ser, respectivamente, las mismas que en APE. Las diferencias significativas para la prueba de Wilcoxon, tanto en la valoración Global como en cada aspecto, refleja grandes discrepancias entre las valoraciones obtenidas por un mismo grupo en *Evaluación* en APE y MAPE (Tabla 5), lo que indica grandes cambios en todos los apartados.

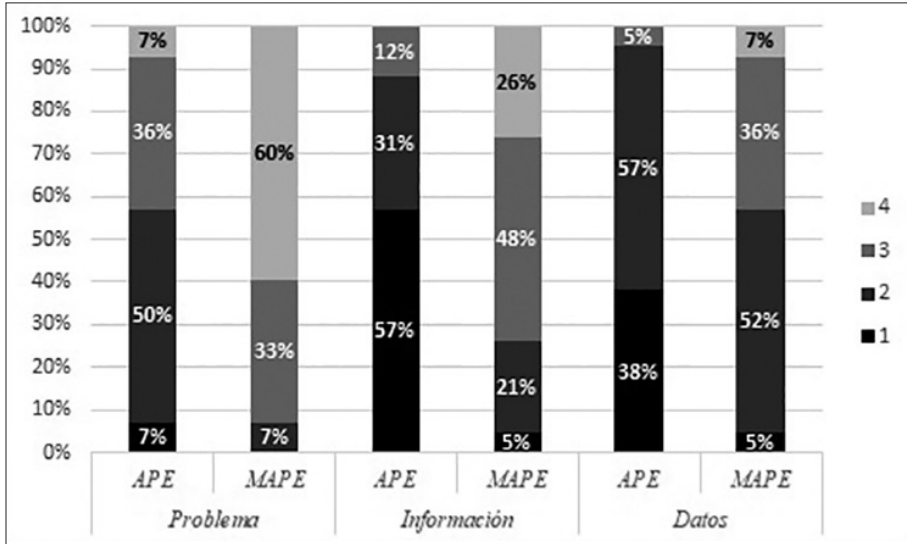
Tabla 5. Correlación de Spearman y prueba de Wilcoxon para *Evaluación* de APE y MAPE

	<i>Coefficiente de Correlación (sig.)</i>	<i>p de Wilcoxon</i>
<i>Problema</i>	0,369 (0,016*)	0,000**
<i>Información</i>	0,387 (0,011*)	0,000**
<i>Datos</i>	0,219 (0,164)	0,000**
Global	0,411 (0,007**)	0,000**

\*p<0,05 \*\*p<0,01

De hecho, la media de las valoraciones Globales de todos los grupos aumentó en MAPE respecto a APE (pasó de 1,98 a 3,10) y lo mismo ocurrió en cada aspecto (la media de las valoraciones de todos los grupos pasó de 2,43 a 3,52 en *Problema*; de 1,67 a 2,95 en *Información* y de 1,98 a 3,10 en *Datos*).

Figura 4. Porcentajes de grupos valorados con 1, 2, 3 y 4 para cada uno de los aspectos *Problema*, *Información* y *Datos* en APE y MAPE

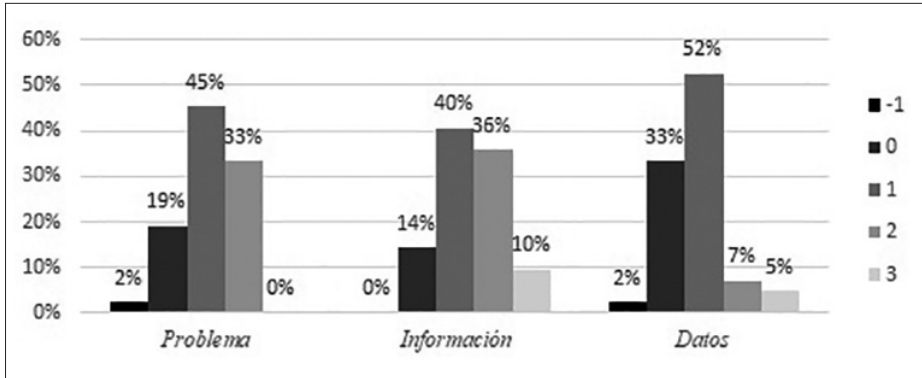


Más concretamente, todos los grupos, salvo dos, mejoraron su valoración Global en el MAPE entre 0,5 y 2,25 puntos respecto a la de sus correspondientes APE. Pero estos cambios no fueron iguales en los tres aspectos considerados. Por ejemplo, el porcentaje de grupos que obtuvo valoraciones mayores que 2 pasó del 43% en APE al 93% en MAPE en *Problema* (es de resaltar que el 60% de los grupos obtuviera una valoración de 4 en MAPE en este aspecto); el porcentaje varió del 12% en APE (ninguno de ellos valorado con 4) al 74% (donde el 26% de los grupos fue valorado con 4) en *Información*, aspecto donde los grupos obtuvieron las valoraciones más bajas en APE; y pasó de 5% en APE al 43% en MAPE (donde el 7% de los grupos fue valorado con 4) en *Datos* (Figura 4).

Profundizando en las diferencias de valoración en cada aspecto, para cada grupo, en todos los casos la diferencia más habitual fue de un punto, si bien en *Datos* las valoraciones variaron de forma diferente a *Problema* e *Información*. La mayoría de los grupos mejoró su valoración en MAPE respecto a APE (78% de los grupos en *Problema*, 86% en *Información* y 64% en *Datos*); además, en todos los aspectos, la mayoría de los grupos mejoró su valoración 1 ó 2 puntos (aunque cabe destacar que algunos grupos consiguieron mejorar su valoración 3 puntos, el 10% de los grupos en *Información* y el 5% en *Datos*); por el contrario, algunos empeoraron su valoración 1 punto (1 grupo en *Problema* y 1 grupo en *Datos*). No hay que olvidar que el comportamiento de *Datos* siempre fue algo diferente en la comparación de

valoraciones, lo que parece indicar que este aspecto supuso mayores dificultades que el resto a un buen número de grupos. Eso puede ser debido a que en ambas entregas fue el aspecto menos trabajado (algunos grupos aún se encontraban diseñando el instrumento de recogida de datos).

Figura 5. Porcentajes de diferencias -1, 0, 1, 2 o 3 en los aspectos *Problema*, *Información* y *Datos* entre las valoraciones de MAPE y APE



Como ejemplo, el grupo que empeoró su valoración en *Problema* y pasó de 4 en APE a 3 en MAPE fue debido a que, en APE, el objetivo del trabajo, «[...] recientes informes publicados afirman que los niños que asisten a colegios bilingües obtienen peores resultados que los matriculados en colegios no bilingües. Queríamos informarnos y comprobar si esos resultados son ciertos y por qué [...]», permitía una justificación basada en las afirmaciones de dichos informes. Sin embargo, en MAPE, el objetivo se convirtió en «[...] valorar las percepciones de una muestra aleatoria de cien ciudadanos de Salamanca sobre la educación bilingüe; atenderemos a su edad, nivel educativo y la fuente de información sobre los colegios bilingües de los encuestados [...]», más amplio y dónde no se justificaba la necesidad de conocer las percepciones de la población sobre la educación bilingüe.

## DISCUSIÓN Y CONCLUSIONES

En este trabajo se analizó la influencia de una intervención educativa basada en el uso de rúbricas de evaluación, en un contexto de evaluación para el aprendizaje como futuros docentes, durante el desarrollo de un Proyecto Estadístico (PE), donde, en un primer momento, los estudiantes, en grupos, se familiarizaron con la rúbrica de evaluación y, posteriormente, la aplicaron al propio Avance del Proyecto Estadístico (APE) y a los de sus compañeros. Estudios desarrollados con la evaluación por pares en las aulas universitarias ponen de manifiesto la necesidad



de formar a los estudiantes en el manejo de rúbricas de evaluación (Ibarra et al., 2012). Esto puede ser más acusado en el caso de la formación de docentes, ya que es un aspecto crucial en su futura profesión.

Los resultados mostraron que el proceso de reflexión desarrollado a partir de la evaluación formativa desarrollada entre iguales y por el profesor fue importante y la reconsideración de fortalezas y debilidades del trabajo realizado tuvo efectos inmediatos dado que las valoraciones de los tres aspectos valorados mejoraron considerablemente en la Mejora del Avance del Proyecto Estadístico (MAPE) respecto a las obtenidas en APE; en concreto, el 78% de los grupos mejoraron sus valoraciones en *Problema*, el 86% en *Información* y el 64% en *Datos*.

El estudio de la consistencia y discrepancias entre las tres formas de valoración, *Autoevaluación*, *Coevaluación* y *Evaluación*, mediante la comparación de la valoración de los APE de cada grupo con la de los compañeros y con la del profesor, mostró que, aunque las tres formas de valoración parecieron consistentes, las valoraciones de los estudiantes, tanto en *Autoevaluación* como en *Coevaluación*, fueron diferentes que la *Evaluación* del profesor; en casi todos los casos los estudiantes fueron más generosos, sobre todo en *Autoevaluación*. Esto podría deberse a la sobrevaloración de estudiantes muy seguros de sí mismos (Lejik y Wyvill, 2001; Sadler y Good, 2006; Wilmot y Crawford, 2005), a la dificultad para manejar la rúbrica de valoración para valorar aspectos en que no se consideran preparados o por estar valorando a colegas (Chang et al., 2012).

Al igual que en los trabajos de Lin et al. (2001a, b) los resultados en *Autoevaluación* superaron en muchos casos a los de *Coevaluación*, donde hubo muy pocas coincidencias; también fueron superiores a los de *Evaluación*, donde las coincidencias superaron en todos los casos el 33%. Si aceptamos que la valoración del profesor fue correcta, esto podría significar una buena comprensión de la rúbrica de valoración por parte de, al menos, la tercera parte de los grupos. Sin embargo, alrededor de un 25% valoraron sus trabajos por encima del profesor con una diferencia de más de 2 puntos sobre 4, que podría indicar laxitud en la aplicación de criterios por parte de los estudiantes, o una gran dificultad o desinterés en la valoración del propio trabajo (Lin et al., 2001a, b; Sadler y Good, 2006). La *Coevaluación* tiene pocas coincidencias con la *Evaluación* y, al igual que en el caso anterior, alrededor de un 25% valoraron sus trabajos por encima del profesor con una diferencia de más de 2 puntos; las razones podrían ser las mismas que para la *Autoevaluación*, pero quizás también puede influir el grado de amistad de las personas del grupo valorado u otros factores personales (Ibarra et al., 2012).

Por otro lado, respecto al tratamiento de cada grupo en su MAPE, fruto de la reflexión y propuestas de mejora formuladas a partir de las valoraciones obtenidas en los APE, los estudiantes, desde el principio se sintieron muy motivados con la realización del PE y se esforzaron en plantear un tema de investigación que

les resultara interesante, que justificaron adecuadamente y para el que plantearon objetivos medibles estadísticamente como muestran los resultados obtenidos en APE. También mostraron interés en la recolección y organización de datos, donde el hecho de participar en la búsqueda o elaboración de encuestas, las dificultades para obtener respuestas en el plazo establecido o qué hacer después con los datos obtenidos provocó una alta concienciación del proceso involucrado en la investigación estadística (Batanero y Díaz, 2011; Burrill y Camden, 2006; Dierker et al., 2016; Franklin, et al. 2005).

De los tres aspectos valorados, el que obtuvo las calificaciones más bajas en APE fue la búsqueda de información (*Información*). Parece que ese aspecto, a pesar de que los estudiantes realizan múltiples trabajos en ese sentido en las asignaturas del Grado, no se consideró importante por los estudiantes ya que, sin ninguna formación en este sentido, fue el aspecto donde se produjo el mayor aumento en las valoraciones obtenidas en MAPE, tanto en la calidad de las fuentes consultadas como en su correcta incorporación en el aspecto de bibliografía. Por tanto, parece conveniente solicitar a los estudiantes para maestro, desde el inicio del Grado y en todas las materias, la realización de búsquedas bibliográficas de calidad, así como la correcta citación de la bibliografía.

En este trabajo se muestra cómo el proceso de evaluación formativa desarrollada entre iguales y por el profesor, diseñado e implementado para la elaboración de un PE, favoreció la reflexión, revisión y mejora del trabajo realizado (Bouzidi y Jaillot, 2009; Chang et al., 2012; Chen, 2010; Miller, 2003; Sadler y Good, 2006). Los diversos aspectos en los que los estudiantes para maestro dieron sentido a la evaluación formativa para su aprendizaje como futuros docentes requeriría más investigación. Sería conveniente buscar más formas de actuación, en las diversas materias relacionadas con la formación didáctico-matemática que cursan los estudiantes para maestro como puede ser la creación de tareas (Cáceres et al., 2015), que favorezcan procesos de reflexión y revisión del propio trabajo, a partir de algún tipo de intervención educativa en la línea de Cáceres et al. (2010).

Los APE y MAPE son avances del Proyecto Estadístico lo que supone, en muchos casos, más una declaración de intenciones que un trabajo terminado. Por ello se considera que, para una mejor toma de decisiones en la elaboración del trabajo final, la rúbrica de valoración debería reestructurarse y organizar diferentes ítems en los tres aspectos considerados. Por ejemplo, se podría separar la justificación de la importancia del problema de estudio de la formulación de objetivos, o la selección de fuentes de documentación de su inclusión en la bibliografía, lo que permitiría profundizar en estudios futuros en las causas concretas que permiten o no la mejora en cada uno de los resultados. Consensuar con los estudiantes para maestro la mejora de la rúbrica de evaluación antes de realizar la sesión permitiría una mejor comprensión de la misma.

Los grupos de estudiantes, durante la intervención educativa, apuntaron propuestas de mejora para su trabajo que el profesor no recogió, lo que no permitió estudiar su consideración en MAPE. Esas intervenciones podrían haber sido objeto de futura investigación en este trabajo u otros con objetivos similares. También sería interesante comprobar el conocimiento profesional para enseñar estadística adquirido con la realización de este trabajo, así como el comportamiento de los grupos en los aspectos de PE que no se trabajaron en esta sesión, así como la secuencia de aula relacionada con su PE.

Esta misma forma de trabajo podría aplicarse a proyectos vinculados a otros conocimientos para la formación de docentes, si bien, requeriría una adaptación de dicha forma de trabajo, así como de los aspectos que deben considerarse en la rúbrica de valoración, que variaría en función de las peculiaridades del contexto formativo, de los profesores o los propios contenidos matemáticos trabajados.

## RECONOCIMIENTOS

Este trabajo se ha desarrollado al amparo de RED8-Educación matemática y formación de profesores (EDU2016-81994-REDT), Proyecto de la Universidad de Salamanca (2017/00111/001), Proyecto Erasmus+ Unión Europea (2017-1-ES01-KA203-038491), Proyecto Ministerio de Economía y Competitividad España (PSI2015-66802-P)

## REFERENCIAS

- Ball, L.D., Thames, H.M. y Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59, 389-407.
- Batanero, C., y Díaz, C. (Eds.), (2011). *Estadística con proyectos*. Departamento de Didáctica de la Matemática. Granada. Recuperado de <http://www.ugr.es/local/batanero/publicaciones%20index.htm>.
- Boud, D., Cohen, R. y Sampson, J. (1999). Peer learning and assessment. *Assessment and Evaluation in Higher Education*, 24(4), 413-426.
- Bouzidi, L. y Jaillet, A. (2009). Can online peer assessment be trusted? *Educational Technology & Society*, 12(4), 257-268.
- Bretones, A. (2008). Participación del alumnado de Educación Superior en su evaluación. *Revista de Educación*, 347, 181-202.
- Browning, C., Goss, J. y Smith, D. (2014). Statistical knowledge for teaching: Elementary preservice teachers. In *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9)*. Recuperado de: [https://iase-web.org/icots/9/proceedings/pdfs/ICOTS9\\_2F2\\_BROWNING.pdf](https://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_2F2_BROWNING.pdf).

- Burrill, G. y Camden (Eds.) (2006). *Curricular development in statistics education: IASE 2004 Roundtable*. Voorburg: International Association for Statistical Education. Recuperado de: [https://iase-web.org/icots/9/proceedings/pdfs/ICOTS9\\_2F2\\_BROWNING.pdf](https://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_2F2_BROWNING.pdf).
- Cáceres, M.J., Chamoso, J.M. y Azcárate, P. (2010). Analysis of the revisions that pre-service teachers of Mathematics make of their own project included in their learning portfolio. *Teaching and Teacher Education*, 26, 1186-1195.
- Cáceres, M.J., Chamoso, J.M. y Cárdenas, J.A. (2015). Situaciones problemáticas auténticas propuestas por estudiantes para maestro. En C. Fernández, M. Molina y N. Planas (Eds.), *Investigación en Educación Matemática XIX* (pp. 201- 210). Alicante: SEIEM.
- Cáceres, M.J. y Chamoso, J.M. (2015). La Evaluación Sobre la Resolución de Problemas de Matemática. En L.J. Blanco, J.A. Cárdenas y A. Caballero (Eds.), *Resolución de Problemas de Matemáticas en la Formación Inicial de Profesores de Primaria* (pp. 225-241). Cáceres: Universidad de Extremadura. Servicio de Publicaciones.
- Carless, D., Joughin, G. y Mok, M.M.C. (2006). Learning-oriented assessment: principles and practice. *Assessment & Evaluation in Higher Education*, 31(4), 395-398.
- Chamoso, J.M. y Cáceres, M.J. (2009). Analysis of the reflections of student-teachers of Mathematics when working with learning portfolios in Spanish university classrooms. *Teaching and Teacher Education*, 25, 198-206.
- Chang, C.C., Tseng, K.H. y Lou, S. J. (2012). A comparative analysis of the consistency and difference among teacher-assessment, student self-assessment and peer-assessment in a Web-based portfolio assessment environment for high school students. *Computers & Education*, 58(1), 303-320.
- Chen, C. (2010). The implementation and evaluation of a mobile self- and peer-assessment system. *Computers & Education*, 55(1), 229-236.
- Cho, K., Schunn, C. y Wilson, R. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, 98(4), 891-901.
- Da Silva, M., Porciúncula, M. y Pinto, S. S. (2014). Teaching Statistics through Learning Projects. *Statistics Education Research Journal*, 13(2), 177-186.
- Dierker, L., Alexander, J., Cooper, J.L., Selya, A., Rose, J. y Dasgupta, N. (2016). Engaging Diverse Students in Statistical Inquiry: A Comparison of Learning Experiences and Outcomes of Under-Represented and Non- Underrepresented Students Enrolled in a Multidisciplinary Project-Based Statistics Course. *International Journal for the Scholarship of Teaching and Learning*, 10(1), Article 2. Recuperado de: <https://digitalcommons.georgiasouthern.edu/cgi/viewcontent.cgi?referer=https://www.google.es/&httpsredir=1&article=1611&context=ij-sotl>.
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M. y Scheaffer, R. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) report: A Pre-K-12 curriculum framework*. Alexandria, VA: American Statistical Association. Recuperado de: [www.amstat.org/Education/gaise/](http://www.amstat.org/Education/gaise/).

- Garfield, J. y Everson, M. (2009). Preparing Teachers of Statistics: A Graduate Course for Future Teachers. *Journal of Statistics Education*, 17(2). Recuperado de: [www.amstat.org/publications/jse/v17n2/garfield.html](http://www.amstat.org/publications/jse/v17n2/garfield.html).
- González, M.T. y Chamoso, J.M. (2015). Enseñanza por proyectos: Una propuesta para la formación de maestros en la educación estadística. *Actas del Congreso La enseñanza de las Matemáticas. Las nuevas metodologías en la enseñanza y el aprendizaje de las Matemáticas Academia de Artillería de Segovia*. (pp. 71-87). Junta de Castilla y León y Asociación Castellano y Leonesa de Educación Matemática Miguel de Guzmán
- Ibarra, M.S., Rodríguez, G. y Gómez, M.A. (2012). La evaluación entre iguales: beneficios y estrategias para su práctica en la universidad, *Revista de Educación*, 359, 206-231.
- Jönsson, A. y Panadero, E. (2017). The use and design of rubrics to support assessment for learning. In D. Carless, S. M. Bridges, C. K.Y. Chan, R. Glofcheski (Eds.), *Scaling up Assessment for Learning in Higher Education* (pp. 99-111). Singapore: Springer.
- Keppell, M., Au, E., Ma, A. y Chan, C. (2006). Peer learning and learning-oriented assessment in technology-enhanced environments. *Assessment & Evaluation in Higher Education*, 31(4), 453-464.
- Knowles, M.S., Holton III, E.F. y Swanson, R.A. (2005). *The adult learner* (6th Ed.). Burlington: Elsevier.
- Lejik, M. y Wyvill, M. (2001). The effect of the inclusion of self-assessment with peer assessment of contributions to a group project: A quantitative study of secret and agreed assessments. *Assessment and Evaluation in Higher Education*, 26(6), 551-561.
- Lin, S.S.J., Liu, E.Z.F. y Yuan, S.M. (2001a). Web-based peer assessment: attitude and achievement. *IEEE Transactions on Education*, 44(2), 13 pp.
- Lin, S.S.J., Liu, E.Z.F. y Yuan, S.M. (2001b). Web-based peer assessment: Feedback for students with various thinking-styles. *Journal of Computer Assisted Learning*, 17(4), 420-432.
- Miedijensky, S. y Tal, T. (2016). Reflection and assessment for learning in science enrichment courses for the gifted. *Studies in Educational Evaluation*, 50, 1-13.
- Miller, P. J. (2003). The effect of scoring criteria specificity on peer and self-assessment. *Assessment & Evaluation in Higher Education*, 28(4), 383-394.
- Nitko, A. J. (2001). *Educational assessment of students*. Upper Saddle River, NJ: Merrill.
- Sadler, P. y Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11(1), 1-31.
- Sung, Y.T., Chang, K.E., Chiou, S.K. y Hou, H.T. (2005). The design and application of a web-based self and peer-assessment system. *Computers & Education*, 45, 187-202.
- Tsai, C.C. y Liang, J.C. (2009). The development of science activities via on-line peer assessment: the role of scientific epistemological views. *Instructional Science*, 37, 293-310.
- Tseng, S.C. y Tsai, C.C. (2007). On-line peer assessment and the role of the peer feedback: a study of high school computer course. *Computers & Education*, 49(4), 1161-1174.

- Watson, J.M. (2002). Doing research in statistics education: More than just data. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a Statistically Literate Society*. Voorburg, The Netherlands: International Statistical Institute. Recuperado de: [https://iase-web.org/documents/papers/icots6/03\\_wa.pdf](https://iase-web.org/documents/papers/icots6/03_wa.pdf).
- Williams, S. (2017). Investigating the allocation and corroboration of individual grades for project-based learning. *Studies in Educational Evaluation*, 53, 1-9.
- Wilmot, P. y Crawford, A. (2005). *Validating the assessment of individuals within undergraduate teams*. International conference on engineering education. Gliwice, Poland. Recuperado de: [https://webpa.lboro.ac.uk/tutors/support/why\\_use/icee2005\\_pa\\_paper.pdf](https://webpa.lboro.ac.uk/tutors/support/why_use/icee2005_pa_paper.pdf).