

Participación del alumnado de Educación Superior en su evaluación

Participation of the High Education students in their assessment

Antonio Bretones Román

*Universidad Complutense de Madrid. Facultad de Educación. Departamento de Didáctica y Organización Escolar.
Madrid, España*

Resumen

El propósito de este trabajo es identificar en qué medida la participación de los estudiantes de Educación superior en su evaluación resulta objetiva y precisa, así como los factores o variables implicados. Este tema -poco investigado en España- aporta una panorámica de la amplitud, naturaleza y conclusiones de la investigación empírica acerca de las formas básicas de dicha evaluación participativa.

El método ha consistido en sintetizar críticamente las revisiones publicadas de esta temática entre 1989 y 2000. Presentamos cinco revisiones bajo tres formas de participación: autoevaluaciones de los estudiantes, evaluaciones de pares y co-evaluación o evaluación compartida entre estudiantes y profesorado.

Exponemos los siguientes aspectos de cada una de las revisiones: datos de identificación, propósito u objetivos, procedimiento de investigación, resultados, conclusiones principales y una crítica personal. Finalizamos con unas conclusiones generales referidas a cuestiones estructurales, validez de los resultados, condiciones de éxito, efectos en el aprendizaje, recomendaciones al profesorado y necesidades para la investigación futura. De estas conclusiones destacamos que las formas de evaluación participativa del alumnado son tan válidas como las del profesorado -siempre que cumplan algunas condiciones- y tienen efectos positivos en el aprendizaje.

Palabras clave: Educación Superior, autoevaluación, evaluación de pares, coevaluación o evaluación colaborativa.

Abstract

The purpose of this study is to identify to what extent the participation of students in Higher Education in the assessment of their own learning can be objective and accurate, as well as to study the factors or variables involved in the process. This topic, not sufficiently researched in Spain, contribute to a panoramic of the amplitude, nature and conclusions of the empirical research on the basic ways of the participative assessment.

The method has been based on critically synthesizing the revisions that were published on this area from 1989 to 2000. Five reviews, involving three different types of assessment participation, are presented here: student self-evaluation, peer assessment and student-teacher joint evaluation or shared assessment.

The following items are included in each review: identification, objectives, research procedure, results, main conclusions and personal review. We end up presenting some general conclusions regarding structural issues, validity of results, conditions for success, influence on the learning, recommendations for the teachers and suggestions for future research. From our conclusions, we highlight that, provided that some conditions are fulfilled, the assessment by students of their learning is as valid as the one made by teachers, with an additional positive contribution to learning.

Key Words: Higher Education, self-assessment, peer assessment, co-assessment or collaborative evaluation.

Antecedentes y fundamentación teórica

La práctica de la evaluación participativa es un recurso apropiado para el desarrollo de algunas competencias –el trabajo en equipo y la toma de decisiones, autocrítica y aceptación de las ideas de los demás, negociación y compromiso para cumplir lo acordado– sugeridas bajo las directrices de la reforma de la enseñanza universitaria, asumidas desde las propuestas del Espacio Europeo de Educación Superior (EEES) a favor de una mayor calidad y protagonismo del alumnado. Las reflexiones de Gimeno (2005, p. 161-180) van en esta dirección enriquecedora de dicha reforma, siempre que se consideren los riesgos y se apoyen las condiciones necesarias como sería la evaluación participativa.

Siguiendo a Somervell (1993) consideramos que cambiar las estructuras de la evaluación en Educación superior exige una enseñanza centrada en el alumnado, la cual responde mejor a las necesidades de éste, ya que lo hace más consciente de sus

potencialidades, debilidades y de su progreso. Igualmente, la evaluación participativa es una parte integral del proceso de enseñanza y aprendizaje que contribuye a que éste sea más profundo.

La participación en evaluación, según Somervell (1993), tiene tres formas básicas:

- La autoevaluación, en la que los estudiantes valoran su propio trabajo, es la más radical, rompe las barreras en la relación estudiante-profesor y el alumnado participa en todas las fases del proceso evaluador, permitiéndole su control.
- La evaluación de pares, en ella valoran el trabajo de sus pares, desde la perspectiva formativa desarrolla las capacidades de reflexión y el producto del aprendizaje, pero desde la sumativa es la más cuestionable.
- La evaluación colaborativa entre estudiantes y profesorado, aquí valoran conjuntamente el propio trabajo o el de los pares, se impulsa el diálogo entre ambos y puede darse en distintas fases y grados.

Señalan Boud y Falchikov (1989), Falchikov y Boud (1989), Bataloso (1995), Fernández Sierra (1996), Espadas (1998-99), López Pastor (2000) y Alvarez (2001) que la participación del alumnado en su evaluación tiene ventajas de tipo intelectual (supone más capacidad crítica y de autonomía), de tipo afectivo (más motivación y autorresponsabilidad) y de tipo profesional (ha dado buenos resultados en el desarrollo de destrezas en los campos de la medicina, la enseñanza, la orientación, la ingeniería y el mundo empresarial y comercial).

Igualmente, Reynolds y Trehan (2000) relacionan la participación del alumnado en su evaluación con la pedagogía crítica y que el apoyo de dicha participación favorece procesos de socialización e integración en grupos. Recuerdan también que el papel del alumnado en la pedagogía crítica es reflexionar sobre su propia experiencia para darle sentido e influir en la dirección y contenido de su aprendizaje y compartir la toma de decisiones. Añaden que el lugar de la evaluación es privilegiado para el análisis de las relaciones de poder, siendo en este sentido no una parte más del proceso de enseñanza y aprendizaje sino la más importante, ya que en ella se puede fomentar la colaboración de estudiantes y profesorado en los demás procesos didácticos.

Más recientemente Tan (2004) sostiene que la práctica de la evaluación participativa de los estudiantes es un medio para darles mayor capacidad de decisión frente al poder unilateral del profesorado, sin embargo esto ha sido cuestionado por quienes piensan que puede servir para disciplinar o controlar al alumnado, entre ellos Taras (2001) que argumenta que para que los estudiantes compartan ese poder deben par-

ticipar en la «evaluación que cuenta», o sea en su evaluación formal con efectos sumativos. Reynolds y Trehan (2000) objetan que esa participación del alumnado en su evaluación formal tampoco es suficiente porque en la evaluación participativa en general y en la autoevaluación en particular subyacen varios tipos de poder –soberano o de los individuos, epistemológico o de las instituciones y disciplinar o de los discursos dominantes– que pueden aumentar o socavar la capacidad de decisión de los estudiantes; para evitar este socavamiento resulta necesario que dichas formas de poder sean comprendidas, es decir considerar que la evaluación participativa no es tanto una cuestión de redistribución de poder entre profesores y estudiantes cuanto de autonomía y aprendizaje. Por tanto, el profesorado debe reflexionar sobre sus prácticas autoevaluadoras y la autonomía debe ser vista no como un fin en sí misma sino como un medio para apoyar un mejor aprendizaje.

Además de las tres modalidades básicas de Somervell (1993) –señaladas anteriormente– vamos a incluir de acuerdo con Topping (1998, p. 252) otras características de la autoevaluación, tales como: cuantitativa/sumativa o cualitativa/formativa, suplementaria o sustitutiva de la realizada por el docente, centrada en el proceso o en el producto y realizada dentro o fuera de clase. Estas modalidades multiplican las posibilidades y ayudan a que cada docente encuentre soluciones para mejorar la autoevaluación.

Algunos inconvenientes de la autoevaluación se refieren al riesgo de sobrevalorarse o infravalorarse en el acto de evaluar y la exigencia de un mayor esfuerzo para los estudiantes y a una planificación más cuidadosa para el profesorado. No obstante, es importante reconocer que el criterio de éxito que suele encontrarse en las investigaciones acerca de la autoevaluación del alumnado se refiere a la coincidencia en las puntuaciones/calificaciones entre estudiantes y docentes al evaluar la misma actuación de aprendizaje, y esto es lo que se reconoce como precisión.

En lo que sigue presentamos las síntesis de cinco revisiones¹ –cualitativas y metaanálisis– de investigaciones de campo o empíricas –individuales y originales– publicadas en la segunda mitad del siglo XX sobre la participación de los estudiantes de educación superior en su evaluación. Un trabajo similar a éste es el llevado a cabo por Abrami, Cohen y

¹ Las fuentes consultadas para este trabajo han sido: Biblioteca CIDE del Ministerio de Educación y Ciencia y Biblioteca de la Facultad de Educación/Centro de Formación de Profesorado de la Universidad Complutense de Madrid, así como los índices bibliográficos Higher Education Abstracts, Research into Higher Education Abstracts y Boletín Internacional de Bibliografía sobre Educación (BIBE). En la bibliografía española sólo hemos encontrado dos trabajos de síntesis de investigaciones, Salinas (1997) y Bretones (2002), pero con diferente enfoque de las cinco aquí analizadas, ya que se trata de estudios donde se recogen respectivamente tan sólo tres y cuatro casos descriptivos de experiencias prácticas de participación del alumnado en su evaluación y correspondientes a diferentes niveles o etapas educativas.

D'Apollonia (1988), acerca de la evaluación de la instrucción o del profesorado por los estudiantes en la educación superior, tratando de ofrecer recursos fundamentalmente metodológicos de revisiones cuantitativas o metaanálisis.

Analizamos a continuación las revisiones de Boud y Falchikov (1989), Falchikov y Boud (1989), Topping (1998), Falchikov y Goldfinch (2000) y la de Dochy, Segers y Sluijsmans (1999). Las cinco síntesis de investigaciones las exponemos agrupadas según las tres formas básicas indicadas de evaluación participativa.² De cada grupo presentamos unas tablas en las que señalamos las características fundamentales.

Revisiones de investigaciones sobre la autoevaluación de los/las estudiantes

Revisión cualitativa de Boud y Falchikov

En la revisión crítica de Boud y Falchikov (1989) se incluyen 48 investigaciones referidas a la autoevaluación del alumnado en la educación superior entre 1932 y 1988. En ella se constata la escasa precisión de los datos aportados y se ofrecen criterios para un mayor rigor metodológico en sucesivos estudios. Se trataba de estudios de evaluación cuantitativa que comparaban la autoevaluación de los estudiantes con la evaluación del profesor. Se calcularon no sólo las frecuencias con respecto a seis variables de aquellos autores que las indagaron, sino que también se enumeran en torno a una quincena de criterios que los estudios individuales y originales debieran precisar para facilitar una revisión consistente de la muestra seleccionada. Algunos de estos criterios son: el motivo de la investigación, las características de los sujetos, contenido de lo que se evalúa, el tipo y los criterios de evaluación, si éstos fueron negociados o no, así como cuál era la variable independiente y los cálculos o estadísticos que debían incluirse. La revisión en las cinco variables contempladas consistía en

⁽²⁾ La caracterización, problemas y soluciones metodológicos que plantean los trabajos de síntesis investigaciones en general, pueden verse en: la revisión citada de Abrami, Cohen y D'Apollonia, que incluye también una dura crítica a las de tipo narrativo; Slavin (1986 y 1987), quien propone el método denominado «síntesis de la mejor evidencia» para aunar las ventajas de las revisiones cualitativas y cuantitativas, respondiendo en el segundo a las objeciones planteadas por Joyce (1987); Giaconia y Hedges (1992), donde se exponen los cuatro métodos de síntesis más destacados; y Morales (1993), que principalmente describe con detalle los pasos del proceso en el metaanálisis.

computar cuántos de tales estudios identificaron sobrevaloración o infravaloración de los estudiantes. Las variables y resultados analizados fueron siguientes:

- ¿Los estudiantes se sobre o infravaloraban en relación con la evaluación realizada por la plantilla docente? 17 estudios informan que hay sobrevaloración, en 12 predomina la infravaloración y en tres no hay una tendencia clara; concluyen los autores que son las circunstancias diferentes las que llevan a una u otra tendencia.
- ¿Los estudiantes de diferente habilidad tendían a sobre o infravalorarse con respecto a los profesores? 11 estudios que contemplan esta segunda variable sugieren que los estudiantes de alto rendimiento tienden a una autovaloración realista y tal vez a infravalorarse, mientras que estudiantes de bajo rendimiento tienden a sobrevalorarse, estos últimos en mayor medida que los primeros.
- ¿Los estudiantes de diferentes niveles académicos tienden a sobre o infravalorarse? Ocho estudios identificados muestran que los estudiantes de cursos más avanzados o graduados tienden a ser más precisos e incluso a aumentar su infraestimación que los de cursos introductorios.
- ¿Los estudiantes mejoran su capacidad para autoevaluarse con el paso del tiempo o la práctica? En siete estudios que contemplan esta variable no hay resultados claros acerca de la influencia del tiempo o de la práctica en la mejora de la habilidad para autovalorarse.
- ¿Los estudiantes se sobre o infravaloran cuando sus autoevaluaciones son usadas para su evaluación formal? De siete casos en los que la autoevaluación cuenta para la evaluación formal, en cinco de ellos se sobrevaloran y en dos no.
- ¿Hay diferencias de género en la autoevaluación? Se sorprenden los autores que sólo seis investigaciones tuvieran en cuenta esta variable, considerando las diferencias en los resultados inconcluyentes por su disparidad y escasez de estudios.

Las conclusiones generales a las que llegan los autores son: no hay una tendencia clara de los estudiantes a la sobre o infravaloración; unos estudiantes se valoraron de modo similar a los profesores y otros no; hay intentos por comparar la autoevaluación de grupos de estudiantes entre sí (por ejemplo estudiantes más capaces con menos capaces, graduados y no graduados); cuando los estudiantes disponen de una escala para valorarse son relativamente capaces, y lo hacen de modo bastante crítico y similar a los docentes (con una leve tendencia de los más capaces a infravalorarse

y de los menos capaces a sobrevalorarse, lo que no debe sorprender afirman los revisores); son necesarias nuevas investigaciones para saber si influye el tiempo y la práctica en la mejora de la habilidad para autoevaluarse; en cuanto a la influencia del género los resultados son contradictorios e inconsistentes, como acabamos de decir.

Por nuestra parte matizar a la revisión de Boud y Falchikov (1989), que utilizando el método de porcentajes facilita una rápida y sencilla panorámica de la situación respecto del objetivo principal de las investigaciones individuales sobre este tema. Estas se centran en determinar en qué medida son precisas las valoraciones que hacen los estudiantes sobre sus propios trabajos y/o aprendizajes en comparación con las valoraciones hechas por el profesorado y en función de qué factores tal precisión puede variar, como la habilidad del alumnado, curso de la carrera, trascendencia en su evaluación formal y el género. No obstante, resulta algo simple y superficial señalar sólo las frecuencias de los estudios originales en una u otra dirección respecto de cada variable contemplada. Destacamos como positivo presentar en orden cronológico los estudios originales y, a veces, los resultados.

Meta-análisis de Falchikov y Boud

Falchikov y Boud (1989), apoyándose en la revisión anterior, realizan esta otra para determinar qué factores o variables son los causantes de que haya consistencia o no en la correspondencia entre la autoevaluación de los estudiantes y la evaluación del profesorado acerca de una misma tarea.

Analizan 57 estudios de educación superior, entre 1961 y 1988, incluyendo casi todos los revisados por Boud y Falchikov (1989). En este trabajo comparan la autoevaluación cuantitativa de los estudiantes con la evaluación del profesor/a. También intentan comprobar si se sobre o infravaloraban, identificando cuatro características (variables independientes): tipos de estudiantes, contenidos (ciencias, humanidades y letras), cursos al que pertenecen, instrumentos y/o criterios empleados, y calidad del diseño (alta o baja). Todas ellas utilizadas con los estadísticos y valores métricos comunes (efectos o variables dependientes): las medias y desviaciones estándar (magnitud o tamaño del efecto), correlaciones y los porcentajes entre las autoevaluaciones de estudiantes y del profesor/profesora. Además comparan esta evaluación alternativa con la tradicional.

Los autores valoran la definición del concepto de autoevaluación y en qué consiste el acuerdo de ésta con la evaluación realizada por los/las docentes. Igualmente

analizan lo que para un mismo grupo o conjunto de autoevaluaciones del alumnado y evaluaciones del docente puede dar lugar a unos mayores o menores porcentajes de acuerdo. Detectan problemas y afirman que podrían solucionarse con procedimientos estadísticos más cuantitativos y complementarios pero no sustitutivos de los cualitativos.

Los 57 estudios seleccionados son distribuidos según la finalidad de cada uno. El grupo principal, con un 33% de trabajos compara la autoevaluación de estudiantes con la evaluación del profesorado. Otro grupo atiende a la mejora de la experiencia de aprendizaje o reflexión de los estudiantes con un 15% de investigaciones y despertar el interés y potenciar la capacidad de autoevaluación es el objetivo de un 13% de estudios. Otro porcentaje también del 13% trataba de ver si la autoevaluación de la actuación variaba según la asignatura, tamaño del grupo-clase y características sociales de éste. Un 11% pretendía comprobar los efectos de los diferentes métodos educativos en la precisión de la autoevaluación; un 6% tenían un punto de vista teórico explícito; un 4% medía las actitudes hacia la autoevaluación; y un porcentaje aún menor de estudios cuantitativos tenían un carácter longitudinal para identificar los efectos de la práctica de la autoevaluación.

El tamaño del efecto fue positivo en la mayor parte de los estudios, es decir la valoración de los estudiantes era superior a la de los/as profesores/as y referido a la mitad de la muestra, cerca de 30 estudios. La mitad de las correlaciones son bajas frente a la otra mitad que son medias y altas (sólo en dos o tres casos es negativa y muy baja). Más exactamente, la correlación media es 0.39 (que para el tamaño del efecto es media alta); en puntuación *z* se corresponde a 0.47, lo que supone que el 68% de las calificaciones de autoevaluación de los estudiantes sea superior a las del profesorado.

En cuanto a los porcentajes del acuerdo entre las autoevaluaciones de los estudiantes y las del profesorado fue más del 70% en la mitad de los estudios que lo calcularon (una quincena), más del 50% en la cuarta parte de la muestra y menos del 50% en la otra cuarta parte (y nos advierten que hay variaciones en la definición del acuerdo entre los investigadores).

Se trataba de relacionar esos efectos o resultados (tamaño del efecto, coeficientes de correlación y porcentajes, en su mayor parte favorables a los estudiantes y con un buen grado de coincidencia entre las autoevaluaciones de los estudiantes y las del profesorado) con las características de los estudios (variables independientes), que son: calidad del diseño (alta o baja), características del estudiantado (nivel del curso, edad o experiencia), naturaleza de los criterios de valoración (global o explícitos, negociados o no), áreas de conocimiento y tareas (ciencias, humanidades y artes; ensayos escritos, presentaciones orales, participación en clase, destrezas profesionales) y tipos de instrumentos empleados (escalas familiares, por ejemplo de 0-10 puntos, o menos familiares). Las hipótesis se constataron en al menos tres de los cinco grupos de variables dependientes:

- El acuerdo, la objetividad o el éxito en la autoevaluación de los estudiantes (comparado con las evaluaciones del profesor) era mayor (o estaba relacionado positivamente) con la calidad alta del diseño, pero cuando no se contemplaba entre los criterios de evaluación el esfuerzo de los estudiantes.
- Parece haber mayor precisión en la autoevaluación en los estudiantes de cursos avanzados (a partir de tercero), pero es más determinante que la edad o el curso la pericia en un campo determinado.
- Los efectos de la materia en dicha precisión, excluida las artes por la escasez de estudios, era mayor en ciencias que en humanidades, de otro lado comparando las áreas profesionales y las académicas así como las de producto (por ejemplo, ensayo escrito) y las de proceso (participación del grupo) son pocos casos y los tres estadísticos comunes son muy dispares entre sí, no obstante aparece la mayor precisión en las tareas profesionales (especialmente en las de enseñanza) con respecto a las académicas, en las de producto con respecto a las de proceso (sobre todo en los ensayos escritos y presentaciones orales).
- No están claras las relaciones o influencia de los criterios en la precisión de la autoevaluación de los estudiantes, además de ser pocos los estudios para cada uno de los siete criterios contemplados parece haber una relativa precisión en los tres criterios donde están implicados los estudiantes (criterios negociados, criterios de los estudiantes y justificación o autoevaluación razonada) y los buenos resultados no están relacionados con el juicio global.
- También son poco consistentes los efectos de los instrumentos, por la gran variedad de éstos y el gran número de discriminaciones que pueden producir (desde sólo dos -aceptable y no aceptable- a varios cientos); parecen estar relacionadas con una buena precisión las escalas de alcance medio, en torno a 100 discriminaciones.

De las conclusiones de Falchikov y Boud (1989) destacamos:

- El valor de la puesta en práctica de la autoevaluación. Aunque presente problemas hay que idear estrategias para que esta capacidad pueda desarrollarse en el alumnado.
- La relativización de la importancia del acuerdo entre la evaluación de los estudiantes y de los profesores, ya que la práctica de la autoevaluación tiene otros valores educativos (motivación, mayor aprendizaje).

Finalmente de esta revisión parece clarificador mantener la indicación de los porcentajes o frecuencias a favor o en contra de la finalidad pretendida, junto a las medidas

propia estadística; aunque no están suficientemente claras las razones del alto número de estudios analizados o los criterios de inclusión y exclusión, lo que también ocurre en la anterior síntesis de Boud y Falchikov (1989).

En la Tabla I se muestran las principales características de los dos estudios anteriores.

TABLA I. Revisiones referidas a la Autoevaluación

	Nº	Periodo	Propósito	Procedimiento	Resultados	Conclusiones
Boud y Falchikov (1989)	48	De 1932 a 1988	Comparar la autoevaluación cuantitativa del alumnado con la evaluación del profesorado de la misma tarea.	Cómputo de infravaloración o sobrevaloración en relación a seis variables.	Elevados porcentajes de investigaciones donde estudiantes se sobrevaloran y se infravaloran. - Más precisión en estudiantes de mejor rendimiento y cursos avanzados. - Mayor sobrevaloración si la autoevaluación forma parte de la evaluación formal.	- No hay clara tendencia a sobrevalorarse o a infravalorarse. - Precisión cuando hay escala de criterios. - Más investigaciones que determinen el valor del tiempo y la práctica, así como sobre las diferencias de género.
Falchikov y Boud (1989)	57	De 1961 a 1988	Identificar los factores determinantes de la consistencia entre las valoraciones de los estudiantes y del profesorado.	Cálculo del tamaño del efecto, correlaciones y porcentajes	-El 68% de sobrevaloración y porcentaje similar de acuerdo con el profesorado. - En cursos avanzados y en las áreas de ciencias más precisión que en cursos introductorios y en los de humanidades. - No están claras las influencias de los criterios ni de los instrumentos usados.	- Las autoevaluaciones pueden mejorarse, para ello es necesario idear estrategias adecuadas. - Relativización de la importancia del acuerdo entre las autoevaluaciones de estudiantes y evaluación de profesores.

Revisiones de investigaciones sobre evaluación de pares

Revisión cualitativa de Topping

La revisión cualitativa de Topping (1998) sobre evaluación entre pares, completa las dos revisiones anteriores al ocuparse de la modalidad de participación del alumnado en la evaluación de su aprendizaje entre iguales. Comprende 109 trabajos, realizados entre 1980 y 1996, 42 artículos (38%) considerados puramente descriptivos y 67

(62%) reúnen datos ordenados de investigación. El autor responde a dos cuestiones: qué tipos de estudios de evaluación de pares se realizan en la enseñanza superior y cuáles son sus efectos en el estudiantado.

Define la evaluación de pares como la medida o juicio que emiten unos estudiantes considerando la cantidad, nivel y calidad de los resultados de aprendizaje de sus pares. Establece una tipología muy exhaustiva, basada en 17 criterios que presentamos en el Cuadro I.

CUADRO I. Revisiones referidas a la Autoevaluación

Variable	Amplitud de la variación
1 Área/materia del currículo	Todas
2 Objetivos	¿De la plantilla o de los estudiantes?
3 Enfoque o propósito general	¿Cuantitativo/sumativo o cualitativo/formativo o ambos?
4 Producto/tarea	¿Tests/puntuaciones/calificaciones o ensayo escrito o presentaciones orales u otras destrezas?
5 Relación con la evaluación de la plantilla	¿Sustitutiva o suplementaria?
6 Peso oficial.	¿Contribuir a la calificación final del evaluado o no?
7 Direccionalidad	¿Unidireccional, recíproca o mutua?
8 Privacidad	¿Anónimo/confidencial/público?
9 Contacto	¿Distancia o cara?
10 Año	¿El mismo o diferente año de estudio?
11 Habilidad	¿La misma o diferente habilidad?
12 Constelación de evaluadores	¿Individuales o parejas o grupos?
13 Constelación evaluada	¿Individuales o parejas o grupos?
14 Lugar	¿Dentro o fuera de clase?
15 Tiempo	¿Tiempo de clase/tiempo libre/informalmente?
16 Requisito	¿Obligatoria o voluntaria para evaluadores/dos?
17 Recompensa	¿Mérito del curso u otros incentivos o refuerzo para la participación?

Además, el análisis y los resultados de todos los estudios son agrupados por el tipo de tarea que evalúan los estudiantes a sus pares: a través de tests y puntuaciones; destrezas de presentación oral; ensayos escritos; trabajos de grupo y proyectos; destrezas profesionales referidas a la medicina, la enseñanza, la orientación, la empresa, la administración y el comercio; evaluación asistida por ordenador.

En esta amplia revisión se encuentra que hay mayoría de estudios con fiabilidad y validez adecuadas en una gran variedad de aplicaciones, esto quiere decir que las evaluaciones de pares coinciden bastante con las realizadas por otros estudiantes (fiabilidad) y por el profesorado (validez). Concretamente, la correlación estadística en 31 de los estudios que se ocuparon de esta cuestión era alta en 18, había gran variabilidad en siete y era baja en seis.

Por otro lado, constata Topping (1998) que cuando la evaluación de pares es usada con tests y adjudicación de puntuaciones los estudiantes la encuentran exigente pero les reduce la ansiedad, produciéndoles ciertos beneficios en su aprendizaje; la evaluación de pares realizada mediante información detallada es asociada con la mejora en destrezas de presentación y valoración; un número relativamente amplio de estudios de calidad en la evaluación de pares en ensayos escritos sugiere que los resultados son tan buenos o mejores que en la evaluación del profesorado; en trabajos de grupo o proyectos y en destrezas profesionales los resultados son respectivamente positivos y de adecuada fiabilidad, aunque son pocos los estudios y suelen estar basados en la percepción de los estudiantes; finalmente la evaluación de pares asistida por ordenador es un sector emergente y pocos los estudios realizados.

También aporta la revisión de Topping (1998) un resumen de los factores que sustentan el éxito o calidad de la evaluación de pares, entre los que destacamos:

- Clarificar los objetivos y las características de los estudiantes y del proyecto de investigación.
- aconsejar que la evaluación entre pares sea por asignación al azar y no por amistad, unidireccional y no recíproca, por grupos y no individual.
- Reclamar que se hagan explícitos los criterios, discutiéndolos y acordándolos entre docentes y estudiantes, proporcionándoles algunas actividades de entrenamiento previo a su aplicación.

Concluye Topping (1998) que la evaluación de pares alcanzará una mayor expansión en la siguiente década y que los informes sobre las prácticas variadas de ella deberían ser precisos en su caracterización, facilitando así subsiguientes revisiones cualitativas o cuantitativas.

Por nuestra parte, añadir que el conjunto de factores de éxito o rasgos de alta calidad del diseño en la implementación de la evaluación de pares señalados por Topping está en la misma línea y completa las dos revisiones anteriores. Además, consideramos positivo de esta síntesis la gran amplitud de la muestra de estudios primarios analizados, la tipología exhaustiva elaborada a partir de ellos y la clarificadora distribución de los resultados según las tareas. Sin embargo hay una cierta dispersión en los objetivos y en la descripción del método de análisis; la breve comparación que hace de la fiabilidad entre la evaluaciones de pares y las autoevaluaciones resulta algo prolija y equívoca, en todo caso no está claro de ambas formas de evaluación participativa cuál es más consistente. Por último, tras reconocer que muchas investigaciones primarias

confunden fiabilidad con validez en la evaluación de pares, decide estudiar sólo la validez, pero continúa utilizando indistintamente ambos términos.

Metaanálisis de Falchikov y Goldfinch

El metaanálisis de Falchikov y Goldfinch (2000), recoge 48 estudios sobre evaluación de pares en la Educación Superior, referidos a distintas áreas de conocimiento, entre los años 1959 y 1999. El propósito es determinar la fiabilidad y validez de las valoraciones hechas por los estudiantes de las tareas realizadas por sus compañeros/as. De manera semejante a la revisión anterior de Topping, estas autoras consideran que muchos estudios primarios investigan el grado de acuerdo entre las valoraciones emitidas por unos y otros pares (fiabilidad), pero como al profesorado le preocupa más el acuerdo entre las valoraciones de los estudiantes y las de los docentes (validez) es por lo que se ocuparán en su revisión de ésta y no de aquella.

Las autoras hacen referencia a los trabajos presentados anteriormente por Falchikov y Boud (1989) y Topping (1998), indicando brevemente algunas semejanzas de fondo y forma, procediendo del mismo modo. Además atienden al análisis de algunos factores que en su opinión habían sido tratados superficialmente en aquéllos, como las variaciones en función de la tipología de las materias (Ciencias, Ciencias Sociales, Medicina, Ingeniería, Artes) y el tipo de agrupamiento de los estudiantes evaluadores (individual, en parejas o en grupos).

Falchikov y Goldfinch (2000) explican cómo llevaron a cabo el metaanálisis, informando de las fuentes o bases de datos consultadas, mediante una serie de palabras claves relacionadas con el tema para seleccionar las investigaciones originales cuantitativas. Las variables independientes consideradas fueron: identificadores del estudio, características de la población estudiantil, área de conocimiento, nivel del curso, criterios de valoración, calidad del diseño y número de evaluadores implicados; y las variables dependientes, a saber, tamaño del efecto o índice estándar de desviación, correlación estadística y porcentajes de correspondencia entre las valoraciones de los estudiantes y del profesor.³

³ Explican también cómo determinaron el índice de calidad para cada estudio individual. El diseño era valorado pobre si se daban al menos cuatro de estos siete criterios: inadecuada información sobre las características de la población, igualmente sobre las características de otro estudio, tamaño pequeño de la muestra e inadecuadas generalizaciones, inapropiadas tareas requeridas a los estudiantes, no proporcionarles criterios y estructura, uso inapropiado de los criterios e inadecuada información procedimental que impida su replicación.

En relación con dichas variables independientes y dependientes, Falchikov y Goldfich (2000, p. 315) establecieron los factores que eran determinantes de mayor grado de acuerdo entre las evaluaciones de los estudiantes y del profesorado:

- La evaluación de pares que requiere la calificación o puntuación de varias dimensiones individuales parece ser menos válida que el juicio global basado en la comprensión de los criterios.
- La naturaleza de la tarea parece influir en la validez de la evaluación de pares. La valoración de pares de productos y procesos académicos parece corresponder más fielmente a las valoraciones del profesorado.
- Los estudios que son bien diseñados parecen dar mejores acuerdos que aquellos con pobres diseños experimentales.
- No hay evidencia para apoyar la superioridad de las valoraciones de pares múltiples sobre las valoraciones únicas o singulares. Las valoraciones a través de un amplio número de pares (más de 20) parecen conducir a un acuerdo más pobre.
- No hay diferencias claras en validez de las evaluaciones de pares en términos de área disciplinar en las que tienen lugar, pero los pares en materias relacionadas médicamente tienen tendencia a concordar menos en algunos casos.
- La familiaridad y la propiedad de los criterios tiende a mejorar la evaluación de pares.
- La evaluación de pares llevada a cabo en cursos de nivel avanzado no es más válida que la realizada en cursos introductorios, en general.

Concluyen Falchikov y Goldfinch (2000) que su trabajo ofrece información útil acerca de un contexto óptimo para una evaluación de pares exitosa y que convendría investigar en el futuro los efectos de otras variables, tales como la interacción entre variables, la repetición de la experiencia en evaluación de pares, el género y las relaciones de amistad/enemistad de los pares.

Finalmente hemos de destacar de esta revisión metaanalítica la doble comparación que se hace, por un lado con el metaanálisis de Falchikov y Boud (1989) y por otro, con la revisión cualitativa de Topping (1998), explicitando las diferencias principales de fondo y forma con ambas síntesis; pero nos parece insuficiente la argumentación aducida para que se abandone el uso de los porcentajes o recuento de votos como medida o método para determinar la validez en las síntesis de los estudios primarios sobre la evaluación de pares, que no obstante Falchikov y Goldfinch incluyen en su trabajo junto con otros índices estadísticos. En la Tabla II incluimos los datos más señalados de los trabajos de evaluación de pares.

TABLA II. Trabajos de Evaluación de pares

	Nº	Periodo	Propósito	Procedimiento	Resultados	Conclusiones
Topping (1998)	109	De 1980 a 1996	Identificar tipos de evaluación de pares que se realizan en la educación superior y sus efectos en estudiantes.	Analiza los estudios primarios agrupados por tipo de producto o tarea evaluada por los/las estudiantes a sus pares.	<ul style="list-style-type: none"> - Fiabilidad y validez en gran variedad de aplicaciones. - Favorables en el uso de tests y puntuaciones, presentaciones orales y ensayos escritos - No son tan claros en trabajos de grupo o proyectos y destrezas profesionales. - Éxito si se cuida la distribución de evaluadores y evaluados y si participan los estudiantes en los criterios de valoración. 	<ul style="list-style-type: none"> - Tener en cuenta las 17 topologías. - Informar de características de los diseños. - Atender a las facilidades y dificultades del contexto. - Uso de tests, puntuaciones y ensayos escritos.
Falchi-kov y Gold-finch (2000)	48	De 1959 a 1999	<ul style="list-style-type: none"> - Comprobar la fiabilidad y validez de la evaluación de pares. - Refinar los procedimientos anteriores de Falchikov y Boud (1989) y mejorar la revisión de pares de Topping (1998). 	- Relación entre variables independientes (calidad del diseño, área de conocimiento, nivel de estudiantes y criterios de valoración) y v. dependientes (tamaño del efecto, correlaciones y porcentajes).	<ul style="list-style-type: none"> - Mayor acuerdo entre evaluación de pares y profesorado cuando se trata de tareas académicas y se emite un juicio global a partir criterios. - No incide significativamente el área de conocimiento ni el nivel del curso. 	<ul style="list-style-type: none"> - Tener en cuenta las 17 topologías. - Informar de características de los diseños. - Atender a las facilidades y dificultades del contexto. - Uso de tests, puntuaciones y ensayos escritos. - Requiere condiciones prácticas para ser exitosa. - Pendiente: estudio de la relación entre variables, influencias de la práctica, género y relaciones de amistad.

Revisión de Dochy, Segers y Sluijsmans sobre la autoevaluación del alumnado, evaluación de pares y co-evaluación en Educación Superior

Dochy, Segers y Sluijsmans (1999), realizan una revisión de 63 estudios a partir de la fuente de datos del ERIC «Educational Resources Information Center», entre 1987 y 1998. Toman como punto de partida la necesidad de cambiar los procedimientos de evaluación, debido a la extensión actual hacia la educación permanente y el cambio en las metas de la Educación Superior, orientadas hacia la formación del pensamiento autónomo, recursos reflexivos y procedimentales para enfrentarse a retos como las nuevas tecnologías.

Consideran la autoevaluación, la evaluación de pares y la co-evaluación como las alternativas a las formas tradicionales de evaluación, centradas en los exámenes escritos

para comprobar el dominio del contenido informativo y conceptual de una materia. Denominan su revisión de narrativa porque la entienden como el análisis cuidadoso de todos y cada uno de los estudios por separado y agrupando las aportaciones en cada modalidad bajo tres perspectivas: definición, resultados principales e implicaciones para los prácticos. Los resultados están referidos a la influencia de habilidades diferentes de los alumnos, el efecto del tiempo, la precisión, el efecto en el aprendizaje, los métodos utilizados y el contenido o materia. Todos ellos ejemplificados con los estudios específicos correspondientes. Trabajan bajo cuatro agrupaciones: autoevaluación; evaluación de pares; autoevaluación y evaluación de pares; y autoevaluación y evaluación de pares relacionadas con la co-evaluación de alumnado y profesorado. Destacan dos tipos de resultados por su tratamiento sistemático en los estudios originales: precisión y efecto en el aprendizaje; por el contrario se presta escasa atención a los instrumentos utilizados, la influencia de las habilidades de los estudiantes y el efecto del tiempo. Con referencia a la precisión en las cuatro agrupaciones las reflexiones más destacadas son que:

- En el ámbito de autoevaluación se recogen dos estudios o investigaciones empíricas individuales, en uno de ellos se encuentra una alta correlación entre la evaluación de los estudiantes y la evaluación del profesor y en el otro hay discrepancias entre ambas.
- En el ámbito de la evaluación de pares se analizan cinco estudios, tres de ellos con alta coincidencia con la evaluación del profesor y dos con baja.
- En el ámbito de la autoevaluación y evaluación de pares se presentan seis estudios, resultando en todos ellos un alto nivel de acuerdo entre ambas.
- En los ámbitos de en autoevaluación, evaluación de pares y coevaluación se analizan siete estudios, habiendo bastante coincidencia en cuatro de ellos.

Por tanto, en general el grado de precisión, de las valoraciones hechas por los estudiantes en los cuatro ámbitos considerados en los 20 estudios analizados es alto, ya que hay coincidencia en 16 de los mismos (80%) y discrepancia en los cuatro restantes (20%).

En lo que se refiere a los efectos en el aprendizaje contemplados en las tres primeras agrupaciones son positivos, al contar con mayor número de investigaciones que los respaldan: la realización de los estudiantes de la evaluación y la calidad del rendimiento en el aprendizaje (diez estudios), el grado de satisfacción de los estudiantes (nueve estudios) y el incremento de la reflexión de los estudiantes sobre su comportamiento (cinco estudios).

En cuanto a las recomendaciones para los prácticos, lo principal es que la aplicación óptima de estas formas de evaluación participativa exige entrenar a los estudiantes, especialmente de cursos inferiores, y usarla preferentemente en cursos superiores; también conviene usarla con finalidad formativa junto a otros procedimientos tradicionales, y que el estudiantado participe en la elaboración de los criterios.

Estos autores llegan a dos bloques de conclusiones: de un lado, que hay bastante evidencia acerca de: las contribuciones de los estudiantes a la valoración son coherentes con las valoraciones del profesorado; que los estudiantes perciben efectos positivos, y que la implicación de los estudiantes es considerada válida, confiable y equitativa, así como contribuidora de la mejora de la competencia. De otro lado, señalan algunas cuestiones que todavía necesitan investigarse:

- En cuanto a los efectos de la participación de los estudiantes en su evaluación sobre los exámenes finales de conocimientos y actuación, aunque parece lógico que mejore la forma de reflexionar sobre el propio pensamiento y actuación, la evidencia de solo dos estudios es insuficiente.
- Por el contrario hay más estudios que tratan de generalizar los efectos de la evaluación de pares y la coevaluación usando diferentes métodos de puntuación y escalas al respecto, pero debería de haber más información sobre el modo de mejorar la calidad de dichas evaluaciones.
- La mayor parte de los estudios analizados por Dochy, Segers y Sluijsmans (1999) no dan suficientes explicaciones sobre los entornos en los que han sido realizados. Aunque estas modalidades de participación en la evaluación parecen intrínsecas a métodos de enseñanza de resolución de problemas y de trabajo en pequeños grupos tutoriales sería interesante comparar con otros marcos, por ejemplo con un entorno centrado en la dirección del profesor.
- El uso de la autoevaluación, evaluación de pares y co-evaluación es coherente con el aprendizaje a lo largo de toda la vida, porque contribuye a reflexionar sobre los procesos de aprendizaje y de conducta. En este sentido sería valioso investigar los efectos a largo plazo de estos métodos de evaluación participativa. La Tabla III muestra las características fundamentales de esta revisión.

TABLA III. Revisión referida a la auto-evaluación, evaluación de pares y co-evaluación

	Nº	Periodo	Propósito	Procedimiento	Resultados	Conclusiones
Segers y Sluijsmans (1999)	63	De 1987 a 1998	- Buscar alternativas a la evaluación tradicional con la evaluación participativa.	Analizan cada estudio bajo la interpretación narrativa y agrupan según modalidad de evaluación: - definición - resultados - implicaciones prácticas.	- En precisión de 20 estudios en 16 de ellos (80%) hay coincidencia entre estudiantes y profesorado y no la hay en cuatro (20%). - En efectos positivos hay mejora en actuación en la evaluación y rendimiento (diez est.), satisfacción de los estudiantes (nueve est.) y capacidad reflexiva (cinco est.).	- Valoraciones de estudiantes coherentes con profesorado y efectos positivos en el aprendizaje. - Nuevos estudios que se ocupen de los efectos de la evaluación participativa, del contexto y los procedimientos.

Destacamos de esta última síntesis de investigaciones la originalidad: al relacionar la evaluación participativa de los estudiantes con la educación permanente en una sociedad que necesita prácticos reflexivos y por ello la recomendación para la investigación futura de estudiar sus efectos a largo plazo, así como el enfoque integrador de las tres formas básicas de evaluación participativa, aunque la coevaluación o evaluación compartida entre estudiantes y profesorado queda algo fragmentada y diluida.

Conclusiones

Las conclusiones más importantes que podemos establecer sobre el tema a partir de las revisiones analizadas las agrupamos en los siguientes apartados:

Referidas a la naturaleza, procedimiento y tipología

Las formas de evaluación participativa de los estudiantes se concretan en tres: autoevaluación, evaluación de pares, y evaluación compartida, realizadas por separado, combinadas entre sí y con las formas de evaluación tradicionales; éstas últimas son predominantes y aquéllas, que se presentan como alternativas, demandan un cambio del proceso de enseñanza y aprendizaje hacia modelos pedagógicos más reflexivos y democráticos.

El procedimiento de selección de las investigaciones de campo en los cinco trabajos de síntesis presentados es similar: estudios identificados en bases de datos estándar, mínimamente precisos y comparables, cualitativos o cuantitativos. El número de investigaciones analizado en cada una de dichas síntesis o revisiones se sitúa en torno al medio centenar, y son conjuntos no del todo excluyentes. Globalmente son un total de doscientos estudios empíricos y correspondientes a la segunda mitad del siglo XX.

La naturaleza y tipología de los estudios son variados, la investigación fragmentada o dispar y una «síntesis de la mejor evidencia» -a pesar de las revisiones existentes- está pendiente de hacerse.

Referidas a la finalidad y objetivos

Los propósitos más generales de estos trabajos son comparar la evaluación de los estudiantes acerca de su aprendizaje con la realizada por el profesorado y sobre el mismo producto o proceso, para determinar su grado de fiabilidad y validez, así como las consecuencias positivas de dicha experiencia en su formación. También buscan ofrecer alternativas a las formas convencionales de enseñanza y evaluación más en sintonía con lo que debe ser hoy la educación, el mundo profesional y la sociedad.

Entre las limitaciones más comunes de estas revisiones destacamos: escaso número de estudios en determinados productos y áreas, la mayor parte de los estudios se basan en la percepción de los estudiantes y hay poca atención a la habilidad del alumnado para realizar estas formas de evaluación participativa. Además, constatamos que en las revisiones sobre la evaluación de pares se abandona el tema de la fiabilidad para centrarse sólo en el de la validez o se identifican ambos conceptos.

Referidas a las conclusiones principales

Gran parte de los estudios concluyen que los estudiantes dentro de ciertas condiciones pueden calificarse/evaluarse con igual o superior precisión que el docente aunque tienen limitaciones personales y técnicas al igual que el profesorado.

Tanto si la evaluación es sumativa -basada en puntuaciones- como formativa -basada en un *feedback* abierto y detallado- se produce una mejora del aprendizaje y mayor satisfacción en el estudiante.

No todos los estudiantes ni en todas las condiciones se evalúan con igual precisión o imparcialidad. Así, dentro de una tendencia general de leve predominio de la sobrevaloración del estudiantado, la mayor precisión se produce en diseños de calidad y usando criterios de valoración negociados.

Los resultados son contradictorios en lo que se refiere al efecto de las variables área disciplinar y nivel de curso de los estudiantes sobre la precisión de la evaluación. En las autoevaluaciones parece registrarse más precisión en ciencias y en estudiantes de los cursos superiores. Sin embargo esta tendencia no se confirma en la evaluación de pares. Por otro lado, en las evaluaciones de pares, pero no en las autoevaluaciones, se registra mayor precisión cuando se usan juicios globales sobre la base de criterios específicos y en tareas académicas más que en las profesionales.

En relación con los productos evaluados por los pares, los resultados más consistentes se encuentran en los tests y puntuaciones, ensayos escritos y presentaciones orales; pero son menos sólidos, por la escasez de estudios realizados, en los trabajos de grupo o proyectos, en destrezas profesionales y en aprendizaje asistido por ordenador.

Recomendaciones para la evaluación participativa

Algunas recomendaciones para el profesorado son: usar aquella tipología que mejor se adapte a las posibilidades y limitaciones de su contexto, así como el producto y proceso que estén más apoyados por la cantidad de estudios realizados, como tests y puntuaciones, presentaciones orales y ensayos escritos; y como hemos dicho, atender a la buena calidad del diseño y a la participación de los estudiantes en la determinación de los criterios.

Las investigaciones futuras deberían ser más precisas en la tipología usada, los factores de calidad, las características del estudiantado y del diseño investigador, describir el contexto, indicar las razones del éxito y sugerir las medidas de mejora de los procedimientos usados.

Entre las necesidades de la investigación en esta temática constatamos las siguientes: esclarecer la influencia de otros factores como el que la autoevaluación y evaluación de pares formen parte de la evaluación formal del alumnado; la evaluación de pares formativa basada en el uso del portafolios o dossier con la inclusión de materiales múltiples; los efectos de la experiencia de los estudiantes y del género; las diferencias entre la autoevaluación de los estudiantes y la evaluación de pares; y la muy escasa presencia de autoevaluación y evaluación de pares sobre la participación de los alumnos/as en clase.

Finalmente añadir que este trabajo, de revisiones cualitativas y cuantitativas de la investigación primaria, aporta información para obtener una visión global de la participación del alumnado de Educación superior en la evaluación de su aprendizaje, un tema tan importante como aún poco cuidado. Su puesta en práctica contribuiría a potenciar tres requisitos de la buena evaluación: que sea motivadora (planteándola como un reto), continua (integrándola más en el proceso de enseñanza y aprendizaje) y formativa (mejorando dicho proceso); así se conectaría con la formación de calidad propuesta en la reforma del Espacio Europeo de Educación Superior.

Referencias bibliográficas

- ÁLVAREZ, J. M. (2001). *Evaluar para conocer, examinar para excluir*. Madrid: Morata.
- ABRAMI, PH.C., COHEN, P. A. Y D'APOLLONIA, S. (1988). Implementation problems in meta-analysis. *Review of Educational Research*, 58 (2), 151-179.
- BATALOSO, J. M. (1995). ¿Es posible una evaluación democrática? O sobre la necesidad de evaluar educativamente. *Aula*, 35, 3-77.
- BOUD, D. Y FALCHIKOV, N. (1989). Quantitative studies of student self-assessment in Higher Education: a critical analysis of findings. *Higher Education*, 18, 529-549.
- BRETONES, A. (2002). Participación del alumnado en la evaluación de sus aprendizajes. *Cooperación Educativa/Kikiriki*, 65, 6-15.
- DOCHY, F., SEGERS, M Y SLUIJSMANS, D. (1999). The use of self, peer and co-assessment in Higher Education: a review. *Studies in Higher Education*, 24 (3), 331-350.
- ESPADAS (1998-99). Evaluación, investigación y práctica docente: un proyecto de innovación curricular. *Cooperación Educativa/Kikiriki*, 51, 64-68.
- FALCHIKOV, N. (1995). Peer feedback marking: developing peer assessment. *Innovations in Education and Training International*, 32 (2), 175-187.
- FALCHIKOV, N Y BOUD, D. (1989). Student self-assessment in higher education: a meta-analysis. *Review of Educational Research*, 59 (4), 395-340.
- FALCHIKOV, N. Y GOLDFINCH, J. (2000). Student peer assessment in Higher Education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70 (3), 287-322.
- FERNÁNDEZ SIERRA, J. (1996). ¿Evaluación? No, gracias, calificación. *Cuadernos de Pedagogía*, 243, 92-97.

- GIACONIA, R. M. Y HEDGES, L. V. (1992). Síntesis de las investigaciones sobre la eficacia de la enseñanza. En T. HUSÉN Y T. N. POSTLETHWAITE (dirs.), *Enciclopedia Internacional de la Educación*, 8, (5204-5222), Madrid/Barcelona: MEC/Vicens-Vives.
- GIMENO, J. (2005). *La educación que aún es posible*. Madrid: Morata
- JOYCE, B. (1987). A rigorous yet delicate touch: a response to Slavin's proposal for «best-evidence» reviews. *Educational Researcher*, 16 (4), 12-14.
- LÓPEZ PASTOR, V. (2000). *Evaluación compartida. Descripción y análisis de experiencias en Primaria, Secundaria y formación del profesorado en el área de Educación Física*. Sevilla: M.C.E.P.
- MORALES, P. (1993). Líneas actuales de investigación en métodos cuantitativos: el meta-análisis o síntesis integradora. *Revista de Educación*, 300, 191-221.
- REYNOLDS, M. Y TREHAN, K. (2000). Assessment: a critical perspective. *Studies in Higher Education*, 25 (3), 267-278.
- SALINAS, D. (1997). La evaluación no es un callejón sin salida. *Cuadernos de Pedagogía*, 259, 44-48.
- SLAVIN, R. E. (1986). Best-evidence synthesis: Alternative to meta-analysis an traditional reviews. *Educational Researcher*, 15 (9), 5-11.
- (1987). Best-evidence synthesis: why less is more. *Educational Researcher*, 16 (4), 15-16.
- SOMERVELL, H. (1993). Issues in assessment, enterprise and Higher education: the case for self-peer and collaborative assessment. *Assessment & Evaluation in Higher Education*, 18 (3), 221-233.
- TAN, K. H. K. (2004). Does student self-assessment empower or discipline students?. *Assessment & Evaluation in Higher Education*, 29 (6), 651-662.
- TARAS, M. (2001). The use of tutor feedback and student self-assessment in summative assessment tasks: towards transparency for students and for tutors. *Assessment & Evaluation in Higher Education*, 26 (6), 605-614.
- TOPPING, K. (1998). Peer assesment between students in colleges and universities. *Review of Educational Research*, 68 (3), 249-276.

Dirección de contacto: Antonio Bretones Román. Universidad Complutense de Madrid. Facultad de Educación. Departamento de Didáctica y Organización Escolar. Edificio «La Almudena». Rector Royo Villanova, s/n. Ciudad Universitaria. 28040, Madrid, España.
E-mail: antoniob@edu.ucm.es