

Vol.21, Nº4 (Septiembre-Diciembre, 2017)

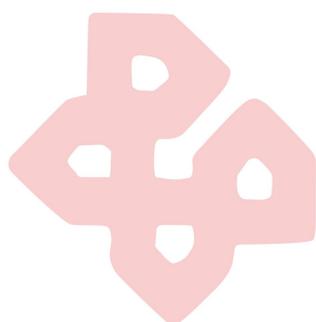
ISSN 1138-414X, ISSNe 1989-639X

Fecha de recepción 25/09/2015

Fecha de aceptación 15/11/2016

LA EVALUACIÓN DE LOS APRENDIZAJES. PROBLEMAS Y SOLUCIONES

Evaluation of the learning. Problems and solutions



José M^a Arribas Estebaranz

Universidad de Valladolid

E-mail: chema2@pdg.uva.es

Resumen:

Las preguntas clásicas acerca de la evaluación: para qué evaluar, qué evaluar, cómo evaluar o quién ha de evaluar los aprendizajes adquiridos por los estudiantes sigue siendo un tema recurrente y polémico en la literatura pedagógica actual que se reaviva periódicamente con motivo de la publicación de las distintas evaluaciones nacionales o internacionales o las polémicas “reválidas” de la LOMCE.

En el ámbito académico coexisten la evaluación formativa y la evaluación certificadora, estableciéndose entre ambas profundas imbricaciones de las que no podemos sustraernos. Ello condiciona, inevitablemente, e incluso polariza todo el proceso de Enseñanza-Aprendizaje. Ambas funciones no son, en absoluto, excluyentes sino complementarias; evaluamos, fundamentalmente, para mejorar; pero, ¿cómo mejorar sin saber de dónde partimos ni adónde hemos llegado? Ciertamente, la mera medición, aislada, descontextualizada, sin consecuencias es un ejercicio estéril que, en el mejor de los casos, solo produce pérdida de recursos y de tiempo, pero también es cierto que no es posible la valoración y consiguiente toma de decisiones en función de esa valoración si no se parte de un conocimiento profundo de aquello que se quiere valorar.

De la calidad de la medición: validez de los indicadores seleccionados, validez y fiabilidad de los instrumentos de evaluación elegidos e idoneidad de las condiciones de aplicación va a depender la calidad de la evaluación y por consiguiente la eficacia de las actuaciones que de ella se deriven.

Así pues, en este artículo intentaremos aportar, justificadamente, tanto reflexiones de carácter teórico como útiles consideraciones de orden práctico que se debieran tener en cuenta a la hora de recoger y valorar los aprendizajes adquiridos por los estudiantes.

Palabras clave: *Calidad educativa, eficiencia de la educación, evaluación por objetivos, evaluación sumativa. medida del rendimiento.*

Abstract:

The classic questions about evaluation: why to evaluate, what to evaluate, how to evaluate, or who is to evaluate the learning acquired by the students are still a recurrent and controversial topic in the current pedagogic literature that periodically revives due to the publication of the different national or international assessments or the controversial "reválidas" of the LOMCE.

In the academic field coexist formative assessment and certification assessment, establishing between the two deep overlapping of which we cannot ignore. This condition inevitably polarizes even entire E-A process. Both functions are not at all exclusive but complementary; we evaluate primarily to improve, but how to improve without knowing where we started from or where we come? Certainly, the mere measurement, isolated, decontextualized, without consequences is a sterile exercise, which in the best case only leads to loss of resources and time, on the other hand it is also true that it is not possible the assessment and subsequent decision making based on that valuation if it is not part of a thorough understanding of what we want to assess.

The quality of the measurement-validity of selected indicators, validity and reliability of assessment instruments chosen and suitability of the conditions of application, will depend on the quality of assessment and therefore the effectiveness of the actions that she derived.

Thus, in this article we will try to provide, justifiably, theoretical thoughts as well as useful practical considerations that should be considered when it comes to collect and evaluate the learning acquired by the students.

Key Words: *Achievement measurement, criterion-referenced evaluation, educational efficiency, quality of education, summative evaluation*

1. Introducción

La evaluación en general y más concretamente, la evaluación en el ámbito educativo es, quizás, uno de los temas más controvertidos -tal vez porque pone en evidencia la idiosincrasia y eficacia personal e institucional- y uno de los más prolíficos de la literatura pedagógica desde hace ya algunas décadas. A pesar de todo, la polémica sigue abierta, tanto en muchos de sus planteamientos teóricos como en su aplicación práctica. La promulgación de la LOMCE, por ejemplo, y su propuesta de evaluación certificadora y externa de los aprendizajes adquiridos por los estudiantes ha reavivado una polémica siempre latente. Hace ya 25 años De Ketele (1988, p. 5) escribía:

La evaluación se reconoce actualmente como uno de los puntos privilegiados para estudiar el proceso de enseñanza-aprendizaje. Abordar el problema de la evaluación supone necesariamente tocar todos los problemas fundamentales de la pedagogía. Cuanto más se penetra en el dominio de la evaluación tanta más conciencia se adquiere del carácter enciclopédico de nuestra ignorancia y más ponemos en cuestión nuestras incertidumbres. Cada interrogante planteada lleva a otras. Cada árbol se enlaza con otro y el bosque aparece como inmenso.

Sin embargo, sigue existiendo, a estas alturas, un alto grado de confusión terminológica e ideológica; en demasiadas ocasiones se utilizan, indiscriminadamente, sin la suficiente precisión, términos tales como valoración, medición, evaluación, calificación, etc. e incluso cuando hablamos de evaluación

cada uno lo hacemos con unas connotaciones muy diferentes en las que subyacen las distintas concepciones ideológicas e incluso la personalidad de cada uno.

Conviene, pues, delimitar claramente los conceptos sobre los que vamos a reflexionar, ya que muchas veces las discrepancias surgen por un problema puramente nominalista: no asociamos a un mismo significante el mismo significado, lo cual, cercena, ya a priori, cualquier atisbo de rigor argumentativo.

1.1. Delimitación conceptual: evaluación continua, formativa... y evaluación puntual, diagnóstica, certificadora...

Por evaluación asumimos la definición que proponen Gairín, Carbonell, Paredes, y Santos, (2009, p. 8): “el proceso de recogida, análisis e interpretación de resultados con el fin de valorarlos y que conlleva una toma de decisiones”.

Partiendo de esta definición y haciendo abstracción de las innumerables definiciones del concepto de evaluación, tres son los elementos fundamentales que la caracterizan: primero, recogida de información -medición-; segundo, valoración, interpretación de esa información y tercero, toma de decisiones en función de los dos primeros.

Distinguiremos pues, las dos funciones primordiales de la evaluación que quedan perfectamente dibujadas en las ya clásicas expresiones de “evaluación para el aprendizaje” -evaluación formativa- y “evaluación del aprendizaje” -evaluación certificadora-

La Evaluación en su dimensión pedagógica, formativa, como un elemento más del proceso de enseñanza aprendizaje se manifiesta más netamente cuando la evaluación no lleva aparejada una calificación o cuando esta no tiene repercusiones más allá de la valoración del progreso del interesado, es decir, cuando únicamente el aprendizaje, *per se*, es el objetivo, bien por la satisfacción que el objeto de ese aprendizaje nos produce en sí mismo -motivación intrínseca- o por las consecuencias derivadas de dicho aprendizaje -motivación extrínseca-. Ejemplos de este tipo de aprendizaje pueden ser la *universidad de la experiencia*, el aprendizaje autodidacta, -cursos de inglés, de guitarra...-, un plan de entrenamiento en un gimnasio, o la preparación para la obtención del carné de conducir en una autoescuela. etc. Han proliferado recientemente multitud de estudios que avalan las bondades de la evaluación formativa, tanto a nivel internacional (Brown y Pickford, 2013; Gauntlett, 2007; Boud y Falchikov, 2007 y Carless, Joughin y Mark, 2006) como a nivel nacional, sobre todo auspiciados por la *Red de evaluación formativa y compartida en docencia universitaria* (Hamodi , López Pastor, A.T. y López Pastor, V. 2014; Fraile, López-Pastor, Castejón y Romero, 2013; Castejón, López-Pastor, Julián y Zaragoza, 2011; Hamodi y López, 2012; Ibarra, Rodríguez y Gómez, 2012; López, Castejón y Pérez, 2012 y Muros y Luis, 2012)

En otras muchas ocasiones, -oposiciones, evaluación Pisa, Evaluación de Diagnóstico, pruebas de acceso a la universidad, al comienzo y al final de un proceso de enseñanza-aprendizaje...- el objetivo de la evaluación es, fundamentalmente

diagnóstico: conocer el punto de partida con el fin de orientar el subsiguiente proceso en una determinada dirección o *certificador*: constatar el grado de consecución de los objetivos previamente determinados con el fin de avalar, clasificar, discriminar... y, también, revisar, rehacer, retomar, reflexionar... en función de los resultados obtenidos.

La evaluación en el ámbito académico, comparte estas dos funciones de la evaluación; la función formativa, pedagógica, suele llevar aparejada, esa otra función, sancionadora, certificadora, habilitadora...; en este caso, de la evaluación se desprende una calificación de la que se suelen derivar importantes consecuencias para el evaluado.

Esta circunstancia condiciona inevitablemente e incluso polariza todo el proceso de enseñanza-aprendizaje.

En ocasiones se pretende contraponer las funciones formativa y certificadora de la evaluación; sin embargo, ambas funciones, no son, en absoluto, excluyentes sino complementarias. Evaluamos, fundamentalmente, para mejorar, pero ¿cómo mejorar sin saber de dónde partimos ni adónde hemos llegado? Medir, interpretar, valorar y tomar decisiones: todo ello constituye el proceso de evaluación, un proceso complejo, global que cobra su verdadero sentido no en una u otra de sus fases sino en la *complitud* de la misma, estableciéndose un bucle interminable, continuamente perfectible.

Ciertamente, la mera medición, aislada, descontextualizada, sin consecuencias es un ejercicio estéril que, en el mejor de los casos, solo produce pérdida de recursos y de tiempo.

Pero también es cierto que no existe evaluación, valoración de la calidad y por consiguiente toma de decisiones en función de esa evaluación si no se parte de un conocimiento profundo de aquello que se quiere valorar. Por lo tanto, la *medición*, entendiendo por tal la constatación de una circunstancia o estado en un momento determinado, al principio o al final de un proceso, repetimos, es imprescindible aunque no suficiente.

La evaluación en su función sumatoria como evaluación de competencias y la evaluación formativa como evaluación para el desarrollo de competencias son dos enfoques complementarios y necesarios de la evaluación de los aprendizajes, que conducen a una concepción global de lo que tiene que ser la evaluación (...) como elemento de formación competencial.» (Villardón, 2006, pp. 61-62, en Gairín, 2008, p.73).

En este sentido, la evaluación de diagnóstico en la etapa de primaria, por ejemplo, tal como recoge la LOE (art. 144) “comprueba el nivel de adquisición de competencias básicas” con el fin de “obtener información para la toma de decisiones sobre la planificación educativa del centro” e “instaurar procesos de mejora en nuestro sistema educativo”, lo que demuestra claramente la subsidiariedad de la

función certificadora de la evaluación respecto de la función formativa, aunque la primera constituya el fundamento de la segunda.

Si la principal preocupación de la evaluación en su dimensión formativa es cómo hacer de ella un instrumento de mejora del aprendizaje integrado dentro del proceso de enseñanza-aprendizaje, la preocupación de la evaluación en su dimensión certificadora en el ámbito académico es precisamente su validez, fiabilidad y practicidad a la hora de constatar la consecución de los objetivos y competencias académicas y profesionales orientada a la mejora.

La mejor conceptualización de ambas definiciones es la que proporcionó en su momento Stake en 1981 en su famosa analogía: Cuando el cocinero prueba la sopa, eso es evaluación formativa; cuando la prueba el cliente, es sumativa-certificadora-.

2. ¿Para qué evaluar?

La evaluación es un elemento indispensable que forma parte natural de cualquier actividad humana -deportiva, académica, laboral...- personal o grupal, que implique un proceso; es la brújula que nos va indicando si llevamos la dirección y velocidad adecuadas hacia nuestro objetivo. Siguiendo con el símil, la evaluación es un corrector automático de dirección, que tiende constantemente a llevarnos a la senda prefijada originalmente, con el fin de corregir errores sobre la marcha, variar la dirección, aumentar la intensidad, dosificar las fuerzas o incluso abandonar a tiempo la carrera, Sin ella andaríamos dando palos de ciego, instalados en la incertidumbre que supone el desconocimiento de nuestra propia situación respecto a la meta. La evaluación, así entendida, es un mecanismo imprescindible de conocimiento y mejora de uno mismo, del grupo y de la actividad que desarrolla.

En el ámbito universitario, por ejemplo, la Introducción a la Guía de Evaluación de Servicios editada por el Consejo de Universidad (2000) considera las razones que explican la necesidad de llevar a cabo la evaluación de la calidad de las universidades. Parece indiscutible que, al margen de otros muchos indicadores -recursos económicos y materiales, instalaciones, tecnología, etc.- el servicio a la sociedad a través de la calidad en la formación inicial -integral- de los estudiantes universitarios, futuros profesionales en breve, es uno de los indicadores fundamentales de la calidad educativa de la universidad y una de sus principales razones de ser.

Tanto las universidades (LOU, art. 2 del Título preliminar) como los centros educativos no universitarios (LOE, art. 120) disponen de autonomía pedagógica y de gestión; pero la libertad va ligada a la responsabilidad. Como cualquier organización o empresa, los centros escolares y las instituciones universitarias han de rendir cuentas de los “procesos de aprendizaje, resultados de los alumnos...” (LOE, art.141) o de la calidad profesional de los titulados (LOU, art. 1 título preliminar) ante la sociedad, máxime en el caso de los centros públicos, financiados con el dinero de todos los contribuyentes.

Así pues, la evaluación en el ámbito académico cumple esa doble función a la que antes aludíamos “la mejora continua y la rendición de cuentas a la sociedad, realizándose esta última a través de la publicación, por ejemplo, de los informes de evaluación” (ANECA, 2003, p. 2).

Por ello es imprescindible tal como manifiestan Dickinson (2000), Sullivan (2002) y Chisholm y Davis (2007) verificar la diferencia existente entre lo que la escuela y la universidad certifican que sus alumnos saben y lo que estos realmente saben sobre todo, en el caso de las universidades, si se pretende avanzar en los procedimientos de acreditación y homologación de los estudios a nivel internacional.

En definitiva, se *mide* para conocer -función diagnóstica y certificadora-, con el fin de evaluar, para mejorar -función formativa-. No se puede mejorar sin conocer y tiene poco sentido conocer si no es para mejorar.

3. ¿Qué evaluar?

La evaluación en función de los objetivos; los objetivos en función del concepto de calidad que propugnemos.

La evaluación de un programa, una institución... ha de hacerse siempre en función de unos objetivos previamente determinados, estos objetivos serán el referente de todo el proceso de evaluación. Así pues, por ejemplo, un club de atletismo de alto rendimiento cumplirá fielmente sus objetivos y será un club de calidad si logra que varios de sus atletas consigan una marca olímpica aunque otros muchos queden frustrados en el camino; por el contrario esos mismos resultados constituirían un fracaso para un club cuyo objetivo fuera el fomento de la actividad deportiva y el deporte de base entre la población escolar.

Evidentemente, los objetivos, los resultados, no pueden conseguirse a cualquier precio, el fin no justifica los medios. Porque pudiera ser, como señala Turull (2010, p. 5) que “se diseñaran políticas que persiguieran únicamente una mejora del rendimiento académico *a cualquier precio*” o por el contrario, “determinadas políticas docentes y académicas podrían estimular la exigencia, el rigor y la calidad de la docencia y no obtener mejores resultados académicos”. Se trata de “incrementar significativamente el rendimiento académico de los estudiantes en un marco de mejora de la calidad docente, de la exigencia y el rigor docentes”.

Ahora bien los objetivos dependerán en gran medida del concepto de calidad que defendamos y este concepto de calidad tiene un componente claramente ideológico, lo que en ocasiones implica que debamos decantarnos por uno u otro modelo.

Al margen de cualquier otro aspecto, la respuesta a la pregunta ¿qué evaluar? Parece, en principio, evidente: hemos de evaluar, el grado de consecución de los

objetivos y el grado de desarrollo de las competencias propuestas, fueren estos cuales fueren. Lo importante es que existan y observen tres condiciones fundamentales:

1. Que estén formulados de forma clara e inequívoca
2. Que sean compartidos, asumidos positivamente por toda la comunidad educativa.
3. Que sean *operativizables*, es decir, que para cada objetivo o competencia se disponga de unos criterios de evaluación que permitan inferir, visibilizar y contrastar el grado de consecución de dicho objetivo o competencia.

No obstante, el asunto es complejo: ¿Por qué unos objetivos y no otros? ¿Quién ha propuesto esos objetivos? ¿Son asumidos y compartidos por toda la comunidad educativa? ¿Los objetivos propuestos a los alumnos han de estar en función de las capacidades y competencias reales que poseen o en función de los objetivos y competencias marcados por ley? Incluso autores como House (1994, p. 31) sostienen que “el evaluador no sólo no tiene que basar su evaluación en los objetivos del programa, sino que debe procurar evitar deliberadamente informarse sobre ellos con el fin de que no le lleven a sesgos tendenciosos”.

En la elección de unos objetivos u otros subyace, casi siempre, un componente ideológico del que se van a derivar muy distintos conceptos de calidad educativa, así como la prioridad que concedamos a unas u otras funciones de la educación, desde una más pragmática y utilitarista: la preparación para el ejercicio de una actividad profesional, hasta otra más humanista y de liderazgo cultural en la sociedad, como reclamaba Ortega y Gasset, ambas contempladas, por ejemplo, en la Ley Orgánica de Universidades (art. 2.1).

No es el momento de entrar en estas, sin duda, interesantes disquisiciones teóricas -vease Barnett (2001). Con el fin de ser operativos vamos a considerar, de entre todos los objetivos posibles y deseables, uno fundamental e indiscutible: la consecución de los objetivos de cada etapa educativa y el desarrollo de las competencias básicas impuestas por ley en los centros docentes no universitarios y las competencias generales y específicas elaboradas por cada universidad y aprobadas por la ANECA.

3.1. La evaluación por competencias: un cambio de perspectiva en cuanto al objeto de evaluación

El modelo basado en competencias, constituye un enfoque interesante y relativamente novedoso ampliamente estudiado (Echeverría, 2001; Mora, 2003; Cano 2005; Marzo, Pedraga y Rivera, 2006; Taberner y Daniel, 2012 y Lupiáñez, 2014) - adoptado tanto en la educación obligatoria (LOE, anexo I) como en la educación universitaria- cuya principal novedad estriba en un cambio de *focalidad*; supone un tránsito desde una concepción parcelada del conocimiento a una concepción integradora. El currículum no gira ya en torno a las asignaturas como compartimentos

estancos, sino que todas ellas, cada una desde su ámbito de conocimiento, coadyuvan a la consecución de los objetivos del título y de las competencias generales y específicas propuestas por cada universidad para ese título.

Existen múltiples definiciones de competencia; una, válida, podría ser esta: “Capacidad de actuar eficazmente en situaciones diversas, complejas e imprevisibles; se apoya en conocimientos, pero también en valores, habilidades, experiencias... (Eurycide, 2002).

Las competencias abarcan, pues, no solo la existencia de determinadas capacidades sino también su puesta en práctica. No se pueden aplicar capacidades que no se poseen pero, sensu contrario, también es cierto que la posesión de esas capacidades es estéril sin la ejecución de las mismas.

Estamos de acuerdo con Perrenoud (2001, en Roca Piera 2008, p.18): cuando dice que “las capacidades son elementos integrantes de las competencias, que deberán ser determinadas y desarrolladas. La capacidad por sí sola o la suma de ellas no definen la totalidad de la competencia. Ésta (la competencia) requiere, además, la llamada *actuación*, es decir el valor añadido que el individuo competente pone en juego o moviliza en un contexto académico y/o profesional preciso (Fernández, 1997; Tejada y Navío, 2005)”.

El aspecto clave del concepto de competencia es la funcionalidad del conocimiento, y la implicación de la persona, del estudiante, en su totalidad: sus aptitudes y actitudes. En el desarrollo de las competencias cobran verdadero significado conceptos tales como aprendizaje significativo, transferencia y generalización del conocimiento, aprendizaje activo, autonomía, autoeficacia, autorregulación, trabajo en equipo, etc.

Por consiguiente, la evaluación de las competencias impone una nueva perspectiva distinta de la tradicional donde impera la comprobación, fundamentalmente del recuerdo de conocimientos, abstractos, sin referentes reales y de carácter fundamentalmente finalista, certificador, poco formativo

Inspirándonos en la taxonomía de de Bloom (1956), en la pirámide de Miller (1990), en los descriptores de Dublín, definidos en el real decreto 1393/2007, en (Rué, J. 2007, p. 72) y en la Ley Orgánica 5/2002 de las Cualificaciones y de la Formación Profesional, proponemos los siguientes niveles en el desarrollo de una competencia:

1. Nivel cognitivo -saber-: tres niveles de procesamiento de la información en función de los requerimientos cognitivos demandados:

1.1. Memorización de información, sin apenas procesamiento de la misma; implica únicamente la memoria, generalmente la memoria a corto plazo.

1.2. Comprensión profunda de esos conceptos; implica reflexión, discriminación, relación, síntesis...

1.3. Basado en el conocimiento y la comprensión, elaboración de un pensamiento creativo, divergente, alternativo... que implica transferencia y generalización. Con frecuencia va más allá de lo cognitivo, insta a la acción.

2. Nivel procedimental -saber hacer-: aplicación de los conocimientos teóricos con oportunidad y eficiencia.

3. Nivel emocional, de personalidad...: curiosidad intelectual, inteligencia emocional, madurez, equilibrio personal, tenacidad, resistencia a la frustración, sentimiento de autoeficacia, habilidades sociales, espíritu de sacrificio, capacidad para trabajar en equipo, liderazgo...

4. Nivel ético: solidaridad, responsabilidad, honestidad, respeto a la diversidad, asunción de valores democráticos...

5. Nivel volitivo: querer hacer.

Quisiéramos hacer notar, en este sentido, cómo los departamentos de selección de personal del mundo empresarial, que suele ir en muchos aspectos por delante del mundo académico -entre otras razones por un sentido puramente pragmático y crematístico- llevan ya bastante tiempo demandando y contratando profesionales que presentan un alto nivel de competencias personales tales como curiosidad intelectual, inteligencia emocional equilibrio personal, tenacidad, resistencia a la frustración, habilidades sociales, espíritu de sacrificio, capacidad para trabajar en equipo, liderazgo... por encima de cualidades meramente académicas o técnicas -fácilmente adquiribles a posteriori y cuya propia empresa se encarga de suministrar-; sin embargo, no es posible dotar de aquellas otras cualidades a quien carece de ellas.

4. Las calificaciones, ¿reflejo fiel de las competencias desarrolladas? Interrogantes acerca de la función certificadora de la evaluación

Como decíamos anteriormente, en el ámbito académico, la función formativa, pedagógica... suele llevar aparejada, esa otra función, sancionadora, certificadora, habilitadora... de la evaluación; en este caso, de la evaluación se desprende una calificación de la que se suelen derivar importantes consecuencias para el evaluado.

Esta circunstancia condiciona inevitablemente e incluso polariza todo el proceso de enseñanza-aprendizaje. Como dice Elliot (1990, p. 35) "[La evaluación] condiciona y artificializa las actividades y procesos de aprendizaje, en virtud de su valor en el intercambio de actuaciones del alumno por calificaciones del profesor".

¿La calificación actúa como un reforzador extrínseco o pervierte la esencia misma del aprendizaje al priorizarse ¿indefectiblemente? el rendimiento frente al aprendizaje, el aprobar frente al saber? ¿Cómo evitar o minimizar este efecto negativo de la calificación?

La calificación es la cuantificación -expresada de forma numérica o verbal, respecto de un referente normativo o criterial- del grado en que un individuo o institución han desarrollado los objetivos y competencias previstos a juicio del/los evaluador/es.

Por lo tanto, las calificaciones, al margen de su imprescindible valor certificador, social, laboral, promocional... adquieren su máxima significación por cuanto son -habrían de ser- la manifestación de la consecución de dichos objetivos y competencias.

A este respecto, tal como recogen Rodríguez y Ruiz (2011, p. 469) son numerosos los autores que ponen en entredicho, por múltiples razones, la validez y fiabilidad de las calificaciones otorgadas por los profesores como medida real del rendimiento.

La *medición* de los aprendizajes adquiridos por los alumnos es una cuestión sumamente compleja por múltiples razones:

En primer lugar, porque, en muchas ocasiones, los objetivos y competencias expresados en los distintos documentos son tan imprecisos y generalistas que admiten bajo un mismo epígrafe, contenidos absolutamente dispares entre sí.

Por otro lado, uno de los principales problemas de la evaluación certificadora es su dificultad a la hora de medir aprendizajes complejos (Williams, 2006). En Álvarez (2008) encontramos una interesante revisión bibliográfica al respecto de donde seleccionamos los siguientes: Elander, Harrington, Norton, Robinson y Reddy (2006) y Webster, Pepper y Jenkins (2000) manifiestan la dificultad de *medir* las competencias comunicativas; Tractenberg, Chaterji y Haramati, (2007), Cowdroy y De Graaff, (2005), manifiestan esa misma dificultad a la hora de *medir* actitudes y Trotter (2006) y Smyth (2004), respecto del trabajo cooperativo.

Por último, porque la calificación se infiere de lo que el alumno manifiesta en las distintas pruebas, cuya valoración depende, en gran medida, de la apreciación subjetiva del evaluador por lo que se precisa contar -hecho que no suele ocurrir- con una serie de indicadores realmente válidos -que midan lo que dicen medir-, “unos estándares -niveles de dominio- establecidos por consenso y unas puntuaciones de corte a partir de las cuales se considera superado, en sus distintos grados, el dominio de la realidad evaluada” (Pérez Juste, 2007, p.15) y unos instrumentos suficientemente fiables -que midan con la suficiente precisión aquello que pretenden medir-.

Las investigaciones (Darby, 2007; Woolf, 2004 y Stowell, 2004) demuestran que no hay entendimiento común sobre el significado de los criterios,

estándares o rúbricas y la consecuente subjetividad con que se califica el aprendizaje, señalándose además que los profesores ocultan o no hacen explícitos a los estudiantes los criterios en los que basan sus juicios sobre la calidad de sus trabajos. (Álvarez, 2008, p. 247).

Las investigaciones ponen de relieve que un mismo evaluador puede asignar valores claramente diferentes -del insuficiente al sobresaliente- a unas mismas pruebas en momentos diferentes y que, por supuesto, este hecho se da claramente cuando los evaluadores son personas distintas. Este hecho pone de relieve una cierta arbitrariedad en las calificaciones, bien porque diferentes evaluadores se fijan en aspectos distintos de la realidad evaluada, bien porque conceden más o menos importancia a determinados componentes de la misma, bien porque unos son más rigurosos que otros al calificar, bien porque... (Pérez Juste, 2007, p. 25)

Así pues, surgen múltiples interrogantes al respecto:

¿Es posible valorar, realmente, las competencias académicas o profesionales o solo ciertas capacidades fundamentalmente cognitivas? ¿Se es capaz de discriminar a los mejores o a los que mejor se adecuan a las demandas de la prueba o del evaluador? ¿Es posible evaluar procesos y competencias complejas a través de pruebas puntuales? ¿Se tienen previstos indicadores de evaluación válidos, fiables y objetivos o dependen en gran medida de la subjetividad del examinador? En la certificación final de una asignatura, etapa o titulación, ¿se ha de valorar el progreso del alumno, en función de sus circunstancias personales, sociales... o la consecución de los objetivos y competencias previstas en dicha asignatura o titulación? ¿Es posible la objetividad sin una evaluación externa? ¿Quién ha de medir? La consecución de dichos objetivos y competencias ¿podría ser constatada por agentes externos al propio proceso de enseñanza-aprendizaje o solo les es posible evaluar en su justa medida a los afectados directamente en dicho proceso -profesor y alumnos?

¿Los resultados, tal como han estudiado algunos autores (Arribas, 2012), muestran suficiente consistencia, independientemente del sistema de evaluación empleado o, por el contrario, son excesivamente dependientes del mismo, hasta el extremo de obtener resultados significativamente distintos en función de aquel?

¿Cómo medir las actitudes, la capacidad de trabajo en equipo, la motivación... aspectos especialmente relevantes en el desempeño de determinadas profesiones? ¿Qué valor predictivo tendrían dichos resultados? ¿Qué uso ha de hacerse de esos resultados?

La respuesta que demos a estas preguntas es una cuestión fundamental. Si llegamos a la conclusión de que se puede medir aquello que se quiere medir, con un suficiente grado de fiabilidad, tendrá sentido analizar, interpretar y tomar decisiones en base a esos resultados; si por el contrario negáramos de antemano esta posibilidad, estaríamos negando la capacidad de intervención, abandonándonos a la supuesta bondad del propio sistema.

Por último, quisiéramos apuntar en esta misma línea, un asunto no menos polémico: el rendimiento académico obtenido por los estudiantes, en tanto en cuanto sea manifestación fidedigna del grado de consecución de esos objetivos y competencias, constituye, a nuestro parecer, un indicador especialmente significativo -entre otros muchos posibles- de la calidad docente del profesor y de la institución. "Cuando se evalúa a un alumno no sólo entra en juego su capacidad de aprender, sino también la capacidad del sistema establecido para enseñarle" (Martín Ciudad, 1991, p. 293).

5. ¿Cómo evaluar?

Una vez asumida la evaluación como un elemento más, inherente al propio proceso de enseñanza-aprendizaje en sus dos funciones principales -*para qué evaluar*- y delimitados claramente los objetivos y competencias básicas -*qué evaluar*-, el *cómo evaluar* puede parecer una cuestión menor y, sin embargo, en la práctica educativa lo más importante no es tanto lo que se hace cuanto cómo se hace aquello que se hace; hasta tal punto que la forma puede condicionar y de hecho condiciona los resultados obtenidos y hasta el propio proceso de aprendizaje (Struyven, Dochy y Janssens, 2005; Birenbaum, 2007 y Burton, 2006).

El cómo estudia un alumno, no es una cuestión trivial, porque el cómo estudia es lo que forma a un alumno (o lo deforma). Un alumno estudiará de memoria o estudiará inteligentemente, procurando entender y relacionar, según sea el tipo de examen o pregunta esperado. El profesor no solo controla qué hace el alumno en clase (escuchar, hacer un ejercicio, etc.) sino que además controla cómo estudia y trabaja en su casa o en otros lugares cuando prepara exámenes y tareas (Morales, 1998, p. 68).

Consideraremos seis aspectos indispensables para una evaluación certificadora de calidad:

Primero. Determinar clara y públicamente el objeto de evaluación, que se constituirá en referente para toda la comunidad educativa.

La evaluación no puede ser un hecho incierto, sorpresivo o caprichoso por parte del evaluador; los objetivos y criterios de evaluación de estos objetivos han de estar perfectamente definidos con el fin de que el alumno tenga el control sobre su propio aprendizaje y por lo tanto sobre su calificación, liberando, de este modo, cuanto sea posible, al profesor de dicha responsabilidad.

Por ello, los objetivos y competencias junto con el resto de los elementos de una programación han de venir reflejados en los decretos y órdenes que regulan las diferentes etapas educativas; en el ámbito universitario ha de plasmarse en las Memorias de las titulaciones de cada universidad así como en las guías didácticas de las diferentes asignaturas, que constituyen, de alguna manera, un documento con "valor contractual" para todas las partes implicadas.

Las guías docentes tienen que reflejar el trabajo de la asignatura y si la evaluación es «una instancia más del aprendizaje», entonces tiene que reflejar todo aquello que está relacionado: criterios, momentos, formas y metodologías. De esta manera, la guía docente como declaración (o contrato) de aprendizaje es una herramienta útil para la (transparencia de la) evaluación”. (Gairin, 2008, p. 74).

Segundo. La existencia de criterios, indicadores de evaluación de los que se infieran claramente esos objetivos que se pretenden evaluar, estableciendo (Pérez Juste, 2007) los estándares -niveles de dominio- y puntuaciones de corte a partir de las cuales se considera superado, en sus distintos grados, el dominio de la realidad evaluada.

Dando por supuesto que cada profesor en su área ha de evaluar los objetivos y competencias prescritos en la guía de su titulación y no otros, ¿los indicadores e instrumentos de evaluación comúnmente utilizados -exámenes, entrega de trabajos, asistencia a clase...- ¿discriminan realmente a quienes han logrado esos objetivos y competencias y en qué grado -metas orientadas al aprendizaje- o a los que mejor se adecuan a las demandas del profesor, midiendo otros aspectos como la capacidad de trabajo, la eficacia, el buen comportamiento, la memoria, la habilidad para “pillar el tranquilo al profesor” -en definitiva, estrategias orientadas más al rendimiento que al aprendizaje?

¿Esos indicadores no deberían ser los mismos para todos?, otro evaluador, ¿obtendría esos mismos resultados con esas mismas pruebas -fiabilidad-?, ¿Cómo explicar la variabilidad de resultados dentro de un mismo grupo de alumnos en las distintas asignaturas o la variabilidad de resultados en una misma asignatura impartida por distintos profesores?

Objetivos/competencias, indicadores de evaluación e instrumentos de evaluación forman una tríada inseparable. ¿Tiene sentido proponer un objetivo que no es evaluable ni/o evaluado? ¿Es verdad, como dice Wert (citado en Moya, 2011) que “en educación como en la mayor parte de las materias, no podemos confiar en aquello que no somos capaces de medir”?

Por otro lado “para que algo sea evaluado como bueno, es preciso establecer previamente criterios de bondad y luego constatar que ese algo está de acuerdo o cumple dichos criterios” (Sadler, 1985, p. 288). Este es, quizás, uno de los problemas más difíciles de solucionar: encontrar dichos indicadores, sobre todo, para aquellos objetivos -de carácter actitudinal- más difícilmente verificables por las dificultades teóricas, técnicas o éticas que entraña su evaluación y calificación.

Entramos de lleno en el verdadero problema de la evaluación certificadora: la articulación de indicadores e instrumentos de evaluación válidos y fiables: ¿Cómo hacer visible, lo invisible? ¿Cómo medir, comparar, jerarquizar... *constructos* tales como la inteligencia, la motivación, las competencias profesionales o simplemente los conocimientos acerca de una materia? ¿Los instrumentos de medida

habitualmente empleados son los adecuados? ¿Se compadece lo que se aprende y cómo se aprende con lo que se mide y cómo se mide?...

Tales *constructos*, al igual que las competencias, no pueden ser medidos directamente, se miden los indicadores que ponen de manifiesto esas competencias; luego es imprescindible, en primer lugar, disponer de unos indicadores de los que se infiera, realmente, la existencia de aquellos.

En esta línea queremos recordar cómo se insiste en diferentes apartados del RD 1393/2007 en la conveniencia de que las competencias propuestas en las nuevas titulaciones universitarias sean evaluables e igualmente la ANECA en sus documentos *Protocolo de evaluación para la verificación de títulos universitarios oficiales* y en la *Guía de apoyo para la elaboración de la memoria para la solicitud de verificación de títulos oficiales* insta a que el modo de expresar las competencias permita la identificación de resultados de aprendizaje que puedan observarse y medirse concretando unos indicadores de evaluación que permitan valorar el grado de desarrollo de cada competencia y su vinculación a actividades de aprendizaje así como a la necesidad de disponer de estrategias que faciliten la evaluación de esos indicadores.

Además, la globalización de la cultura y la economía hace necesario poder comparar productos servicios y cualificaciones académicas y profesionales; una forma de hacerlo es establecer estándares de evaluación tanto a nivel nacional como internacional. En ello ha trabajado la Asociación Internacional para la Evaluación del Rendimiento Educativo, la OCDE por medio del proyecto INES (International Indicators of Educational Systems, el proyecto de Tiana (1988), el proyecto EVA impulsado por el Ministerio de Educación para la evaluación externa de centros escolares, o incluso el proyecto de de Trier y Peschar (1994) sobre indicadores de competencias referidas a objetivos transversales -pensamiento crítico, valores democráticos, destrezas de comunicación...-; sin embargo, sigue siendo aún un asunto pendiente en la práctica educativa.

Por ello, como señala Pérez Yuste (2007, p. 25), en la siguiente cita, que reproducimos, a pesar de su extensión, por lo esclarecedora y pertinente que resulta.

[Es necesario] establecer con claridad las reglas que permiten establecer los niveles que se consideran suficientes porque representan el dominio de la materia evaluada. Y, en ese punto, debemos reconocer que las reglas no existen y que, en consecuencia, deben establecerse por consenso entre personas expertas. Pero deben establecerse, aunque sea difícil, porque, si no fuera así, unos mismos resultados podrían alcanzar diferentes valoraciones lo que, además de injusto para las personas, sería confuso para la sociedad a la que se pretende informar con este tipo de evaluaciones.

La fijación de estándares, esto es, de niveles de dominio claramente definidos que deben ser establecidos como norma, patrón o referencia a lograr en determinados aprendizajes es una condición básica para ordenar la enseñanza

y, por supuesto, la evaluación del aprendizaje. Insisto: la enseñanza, y su evaluación, porque la condición básica para el logro de la validez es la coherencia entre lo evaluado y aquello que debe ser objeto de enseñanza / aprendizaje, siendo este último aspecto el elemento directriz y nunca al contrario.

Por último, hay una cuestión técnica que, también, debe resolverse: la puntuación a partir de la cual se considera que se da el dominio de la realidad evaluada. Una decisión de esta naturaleza puede tomarse por dos procedimientos fundamentales: a partir del juicio de expertos en las diversas materias y en las características del alumnado evaluado, o de forma empírica, estableciendo tales puntuaciones -aunque pueda parecer un contrasentido- a partir de los resultados obtenidos por muestras de sujetos tomados como norma o patrón.

Tercero. La utilización de unos instrumentos de evaluación válidos -que midan lo que queremos medir- y fiables -que lo midan con suficiente precisión-.

Siguiendo a García Llamas (2003) y Camilloni, (1998) la calidad de la medición depende en gran medida de la validez del instrumento: que mida lo que dice medir y no otra cosa, -validez de contenidos, validez predictiva, validez de construcción, validez de convergencia, validez manifiesta, validez de significado, validez de retroacción... - de la fiabilidad: la precisión en la medida que tiene un instrumento independientemente del observador que lo utilice. La fiabilidad implica: estabilidad, exactitud y sensibilidad; es decir, si se volviera a repetir una misma prueba con los mismos alumnos, por el mismo examinador u otro distinto, los resultados tendrían que ser muy parecidos y de la practicidad: la facilidad en cuanto a la administración y la facilidad de análisis e interpretación de los resultados, así como la economía de esfuerzo y coste.

Las competencias básicas y las competencias académicas y profesionales - generales y específicas- objeto de evaluación, son una amalgama compleja de aptitudes, actitudes, motivación... difíciles de medir, por lo que las pruebas de las que se infiera la consecución de esas competencias han de ser lo suficientemente variadas, complejas y precisas para que sean realmente discriminatorias, lo cual supone un verdadero reto a la pretensión de objetividad y practicidad de la evaluación; en este sentido, Burton, (2004, 2005) cuestiona la utilización de las pruebas objetivas para la medición de los aprendizajes adquiridos no tanto por la utilización de dichas pruebas en sí como por la falta de calidad que estas suelen presentar; sin embargo no se descarta, en absoluto, su validez frente a otros tipos de pruebas tal y como señalan Bleske-Rechek, Zeus y Webb (2007). Existen diversos informes de investigación que sugieren cómo controlar los riesgos y superar las limitaciones (Fellenz, 2004, Weber, Frary y Cross, 1995 y Burton y Miller, 1999)

En definitiva, tal como señala Álvarez (2008) las investigaciones divulgadas sobre el valor de las pruebas objetivas continúan afirmando su idoneidad para evaluar el aprendizaje siempre que se cuide la elaboración de los ítems y la calidad

de la prueba aunque en esto radica, precisamente, la dificultad de implementar este tipo de pruebas.

A la vista de todo ello y considerando las aportaciones de diversos autores (Zabalza, 2001, Bolívar, 2007, Gairin, 2008 y Sierra, 2008) proponemos algunas consideraciones de orden práctico:

- Elegir actividades de evaluación, pruebas, que realmente midan aquello que queremos medir que, en términos generales, serán aquellas actividades de evaluación que pongan en juego capacidades cognitivas superiores. Tal como lo expresa Álvarez (2001, p. 115) “Si se pretende desarrollar el pensamiento creativo, crítico, autónomo, formule preguntas que “obliguen” a argumentar de un modo creativo, crítico y autónomo las respuestas”. O dicho más incisivamente:

No espere respuestas geniales de preguntas triviales que no desafían el pensamiento. No espere respuestas brillantes a preguntas que ofenden a la inteligencia. Elabore preguntas que realmente hagan pensar, que estimulen la curiosidad, que espoleen el pensamiento, que activen la imaginación, que provoquen el riesgo reflexivo de buscar respuestas no previstas (Álvarez, 2001, p. 118).

- Coherencia entre los objetivos y competencias que pretendemos desarrollar, las actividades de aprendizaje y los indicadores e instrumentos de evaluación. La *desintonía* entre las competencias demandadas y lo que realmente se enseña y se evalúa en las aulas se produce en todos los niveles educativos como pone de manifiesto Carbonell (2010):

Basta con leer con cierta atención la filosofía que inspira este estudio internacional (PISA) para darse cuenta de que no se miden los aprendizajes concretos de los contenidos tradicionales -lo que se enseña y se aprende en la mayoría de los centros evaluados-, sino las competencias (...). Lamentablemente esos objetivos no figuran en el currículo real de los centros ni en las prácticas pedagógicas de la mayoría del profesorado. Por eso estamos donde estamos (p. 3).

- Evitar que las pruebas de evaluación se conviertan en meros ejercicios puramente academicistas, descontextualizados, sin referentes reales. En este sentido, una buena manera de evaluar el grado de desarrollo de las competencias es por medio de *tareas*; entendemos por *tareas*, situaciones más o menos complejas, referidas a problemas o situaciones reales, o realistas, al menos, cuya resolución requiere la puesta en práctica de todas las capacidades del estudiante o del profesional.
- Diversificar las preguntas o pruebas que midan un mismo indicador: no existe ninguna prueba con validez universal, cada prueba es más o menos idónea en función de aquello que se quiera evaluar, del mismo modo que no existe una

correlación directa entre competencias y pruebas: una prueba puede medir varias competencias y una competencia puede ser medida por medio de varias pruebas.

- Es conveniente evitar la repetición sistemática de un mismo tipo prueba con el fin de evitar el sesgo que supone la “preparación de la prueba”.
- Cuidar la calidad de las pruebas, de modo que sean suficientemente sensibles como para discriminar los distintos grados o niveles en la realización de ese indicador (gradualidad en las demandas).
- Evitar pruebas que impliquen el factor suerte o aquellas excesivamente memorísticas de modo que puedan verse afectadas por el estrés de la prueba.
- Cuidar los condicionantes externos: lugar, hora, cansancio, longitud de la prueba, posibilidad de fraude...
- Ofrecer feed-back siempre que sea posible.

Así pues, en función de estos criterios, la selección de uno u otro instrumento de evaluación no constituye una elección caprichosa, sino que resulta determinante en función de lo que se pretenda medir, adecuándose a los parámetros de validez, fiabilidad y practicidad. Podemos encontrar una extensa lista de referencia, así como la idoneidad de cada uno de ellos en cuanto a su capacidad para poner de manifiesto unas u otras competencias en Márquez, Roca, Solvas, Belmonte, Fernández, y Rodríguez (2011), De Miguel (2004, 2006), Sierra (2008), Gairín (2008) y Montanero (2006).

Cuarto. Practicidad y garantía en las condiciones de aplicación: facilidad en la aplicación y corrección, transparencia del proceso, control del fraude, adaptación a la diversidad...

Quinto. Utilización de los resultados con carácter formativo, no meramente clasificatorio y mucho menos discriminatorio o segregador.

6. ¿Quién ha de evaluar?

El agente evaluador constituye un aspecto especialmente polémico y sensible, que requeriría un espacio del que no disponemos. Véase al respecto (Arribas, 2014). Tan solo apuntar nuestra postura a favor de la necesidad de implementar una evaluación mixta -interna y externa- como garantía imprescindible de calidad educativa; la complementariedad de ambas potencia las ventajas y minimiza los inconvenientes que presentan una y otra por separado-. En este sentido puede considerarse paradigmática y novedosa la experiencia llevada a cabo por Hernández, Samplón, Usón, Letosa, y Mur (2010), en el que se conjugan con éxito ambas formas de evaluación.

En todo caso, sí hemos de señalar al respecto que todos los estudios, casi todos los países y todos los informes al uso -los informes PISA de la OCDE, los informes sobre el panorama de la Educación de la OCDE, los informe del seguimiento de la educación para todos en el mundo de la UNESCO, los estudios internacionales TIMSS y PIRLS del IEA, el informe Eurydice sobre las pruebas nacionales de evaluación del alumnado en Europa, los informes McKinsey, el informe de Hanushek y Woessmann: “The Economics of International Differences in Educational Achievement” o los informes sobre el sistema educativo español de la fundación BBVA-IVIE- contemplan en mayor o menor medida la evaluación externa -junto con la interna- como garantía imprescindible de calidad educativa en sus distintos niveles de concreción curricular: a nivel individual: alumnos, a nivel intermedio: centros y a nivel nacional: sistemas educativos-.

7. Conclusiones

La evaluación sigue siendo un tema controvertido, tanto en sus aspectos teóricos como en su aplicación práctica, como en las consecuencias que de ella se derivan, debido en muchos casos, más a implicaciones ideológicas y políticas subyacentes que a cuestiones puramente pedagógicas. En este artículo hemos querido recoger, reflexionar y aportar algunas propuestas que pudieran concitar el consenso de la mayoría de las voluntades acerca de las numerosas cuestiones que a día de hoy siguen enfrentado las distintas perspectivas existentes a la hora de concebir la educación en general y la evaluación en particular; a saber: la finalidad misma de la evaluación, así como de los contenidos objeto de esa evaluación que oscilaría entre quienes abogan por unos contenidos más utilitaristas, en consonancia con las demandas laborales existentes y quienes abogan por unos contenidos más humanistas, menos pragmáticos e incluso quienes defienden la inconveniencia de la existencia misma de objetivos evaluables; la defensa excluyente de la función formativa de la evaluación en detrimento de su función certificadora como si ambas fueran incompatibles entre sí; la concepción peyorativa de las calificaciones en determinados ámbitos, las dudas sobre su fiabilidad o la posibilidad de inferir o no de ellas la calidad del profesorado o la institución; la reticencia a la evaluación externa de los aprendizajes, así como al uso y publicidad de los resultados obtenidos en dichas evaluaciones

Por otro lado, la incorporación del concepto de competencias ha venido a enriquecer y complicar el proceso de evaluación. Evaluar competencias es mucho más difícil que evaluar únicamente conocimientos o procedimientos, requiere una nueva perspectiva, nuevas formas y nuevos instrumentos. Proponemos finalmente una serie de consideraciones que favorecen un proceso evaluativo de calidad: la evaluación ha de establecerse en función de unos objetivos previamente establecidos, claramente formulados y conocidos por todos los intervinientes en el proceso evaluador; definir una serie de criterios de evaluación válidos -que midan lo que pretenden medir- así como estándares o niveles de aprendizaje y unos instrumentos de evaluación igualmente válidos y fiables -que lo midan con suficiente

precisión-, observando una serie de precauciones con el fin de evitar posibles sesgos - preparación de la prueba, efecto halo, control del fraude...-; ofrecer feed-back siempre que sea posible e instaurar un sistema de evaluación mixto -externa e interna- y por último, utilizar los resultados con carácter formativo, propedéutico no meramente clasificatorio y mucho menos discriminatorio o segregador.

Referencias bibliográficas

- Álvarez, J.M. (2001). *Evaluar para conocer, examinar para excluir*. Madrid: Morata.
- Álvarez, I. (2008). Evaluación del aprendizaje en la Universidad: una mirada retrospectiva y prospectiva desde la divulgación científica. *Revista Electrónica de Investigación Psicoeducativa*, 14,6 (1), 235-272.
- ANECA (2003). *Programa de Evaluación Institucional (PEI) Convocatoria 2003-2005*. Madrid: autor.
- Arribas, J. M^a. (2012). Academic performance in terms of the applied assessment system. *Relieve*, 18, (1), art. 3. Recuperado de http://www.uv.es/RELIEVE/v18n1/RELIEVEv18n1_3eng.htm
- Arribas, J. M^a. (2014). La evaluación externa de los aprendizajes como garantía de calidad educativa. *Avances en Supervisión Educativa*, 22. Recuperado de http://www.adide.org/revista/index.php?option=com_content&task=category§ionid=26&id=185&Itemid=822
- Barnett, R. (2001). *Los límites de la competencia. El conocimiento, la educación superior y la sociedad*. Barcelona: Gedissa.
- Birenbaum, M. (2007). Assessment and instruction preferences and their relationship with test anxiety and learning strategies. *Higher Education*, 32 (2), 749-768.
- Bleske-Rechek, A., Zeug, N., y Webb, R. (2007). Discrepant performance on multiple-choice and short answer assessments and the relation of performance to general scholastic aptitude. *Assessment and Evaluation in Higher Education*, 32 (2), 89-105.
- Boud, D. y Falchikov, N. (2007). *Rethinking assessment in higher education. Learning for the long term*. Londres: Routledge
- Brown, S., y Pickforf, R. (2013). Evaluación de habilidades y competencias en Educación Superior. Madrid: Narcea.
- Burton, R. (2004). Multiple-choice and true/false tests. Reliability measures and some implications of negative marking. *Assessment and Evaluation in Higher Education*, 29 (5), 585-595.

- Burton, R. (2006). Student justice perceptions following assignment. *Assessment and Evaluation in Higher Education*, 31 (6), 655-670.
- Burton, R., y Miller, D. (1999). Statistical modelling of multiple-choice and true/false tests: ways of considering, and of reducing, the uncertainties attributable to guessing. *Assessment and Evaluation in Higher Education*, 24 (4), 399-411.
- Camilloni, A. (1998). *La calidad de los programas de evaluación y de los instrumentos que la integran*. En Camilloni, A., *La evaluación de los aprendizajes en el debate didáctico contemporáneo*. Buenos Aires: Paidós.
- Cano, E. (2005). *Com millorar les competències del docents*. Barcelona: Graó.
- Carbonell, J. (2010). Editorial. *Cuadernos de Pedagogía*, 3, 345.
- Carless, D., Joughin, G. y Mark, M. M. C. (2006). Learning-oriented assessment: principles and practice. *Assessment & Evaluation in Higher Education*, 31 (4), 395-398.
- Castejón, F. J., López, V. M., Julián, J., y Zaragoza, J. (2011). Evaluación formativa y rendimiento académico en la Formación Inicial del Profesorado de Educación Física. *Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte*, 42, 328-346. Recuperado de <http://cdeporte.rediris.es/revista/revista42/artevaluacion163.htm>
- Cowdroy y de Graaff, 2005). Assessing highly-creative ability. *Assessment and Evaluation in Higher Education*, 30 (3), 507-5018.
- Chisholm, C., y Davis, M. (2007). Analysis and evaluation of factors relating to accrediting 100% of prior experiential learning in UK work-based awards. *Assessment and Evaluation in Higher Education*, 32 (1) 45-59.
- De Miguel, M. (2004). *Modalidades de enseñanza centradas en el desarrollo de competencias*. Oviedo: Ediciones de la Universidad de Oviedo.
- Dickinson, M. (2000). An analysis of the accreditation of transferable skills in extra curricular activities within higher education. *Assessment and Evaluation in Higher Education*, 25 (1) 61-70.
- Echevarría, B. (2001). Configuración actual de la profesionalidad. *Letras de Deusto*, 31, 35-55.
- Elander, J., Harrington, K., Norton, L., Robinson, H., & Reddy, P. (2006). Complex skills and academic writing: a review of evidence about the types of learning required to meet core assessment criteria. *Assessment & Evaluation in Higher Education*, 31(1), 71-90.
- ELLIOTT, J. (1990). *La investigación-acción en educación*. Madrid: Morata.

- Fellenz, M. (2004). Using assessment to support higher level learning: the multiple Choice item development assignment. *Assessment and Evaluation in Higher Education*, 29 (6), 703-719.
- Fraile, A., López-Pastor, V.M., Castejón, J. y Romero, R. (2013). La evaluación formativa en docencia universitaria y el rendimiento académico del alumnado. *Aula Abierta* 41 (2), 23-34. Recuperado de [www.uniovi.net/ICE/publicaciones/Aula Abierta/numeros anteriores/i9/05_A_A_Vol.41_n.2.pdf](http://www.uniovi.net/ICE/publicaciones/Aula_Abierta/numeros_anteriores/i9/05_A_A_Vol.41_n.2.pdf)
- Gairín, J., Carbonell, J.L., Paredes, J. y Santos Guerra, M. Á. (2009). *Glosario*. En Paredes, J. (Coord.), De La Herrán A. (Coord.), Santos Guerra, M.Á., Carbonell, J.L., y Gairín, J. (373-378) *La práctica de la innovación educativa*. Madrid: Síntesis.
- Gairín, J., García San Pedro, M^a.J., Gisbert, M., Rodríguez Gómez, D., y Cela, J. M^a. (2008). *La evaluación por competencias en la universidad: posibilidades y limitaciones*. Barcelona: Bellaterra: (Memoria del Estudio) MINISTERIO DE EDUCACIÓN, Dirección General de Política Universitaria Subdirección General de Análisis, Estudios y Profesorado PROGRAMA DE ESTUDIOS Y ANÁLISIS 2008 N^o. Ref. EA2008-0086.
- García Lamas, J. L. (2003). *Métodos en investigación en educación*. (V. II) Madrid: UNED.
- Gauntlett, N. (2007). Literature review on formative assessment in higher education. Middlesex University. Recuperado de: <http://www.mdx.ac.uk/HSSc/cetl/docs/formrevnov07.pdf>
- Ibarra, M. S., Rodríguez, G., y Gómez, M. Á. (2012). La evaluación entre iguales: beneficios y estrategias para su práctica en la universidad. *Revista de Educación*, 359, 206-231. Recuperado de http://www.mecd.gob.es/revista-deeducacion/numeros-revista-educacion/numerosanteriores/2012/re359/re359_11.html
- Hamodi Galán, C., López Pastor, A. T., López Pastor, V. (2014). Red de evaluación formativa y compartida en docencia universitaria: creación, consolidación y líneas de trabajo. *Revista de evaluación educativa*, 3 (1). Recuperado de <file:///C:/Users/Jose%20Maria/Downloads/110-526-1-PB.pdf>
- Hamodi, C., y López, A. T. (2012). La evaluación formativa y compartida en la Formación Inicial del Profesorado desde la perspectiva del alumnado y de los egresados. *Psychology, Society, & Education*, 4(1), 103-116. Recuperado de <http://www.psyse.org/articulos.php?id=87>
- Hernández, J., Samplón, M., Usón, A., Letosa, J., y Mur, J. (2010). Evaluación externa: Un reto posible en nuestra Universidad. Recuperado de www.unizar.es/icee04/innovadoc/15_Eval_Externa_def.pdf

- House, J. (1994). *E. Evaluación, Ética y Poder*, Madrid: La Muralla.
- López, V. M., Castejón, J., y Pérez, Á. (2012). ¿Implicar al alumnado en la Evaluación en la Formación Inicial del Profesorado? *REMIE Multidisciplinary Journal of Education Research*, 2(2), 177-201.
- Lupiañez, J.L. (2014). Competences of the elementary school teacher. *Educacao & Realidade*, 39 (4), 24-36.
- Márquez, J., Roca, J., Solvas, M^a.J., Belmonte, T., Fernández, C. y Rodríguez, D. (2011). Resultados de aprendizaje en los nuevos títulos de grado. *RED-DUSC, Revista de Educación a Distancia. Sección de Docencia Universitaria en la Sociedad del Conocimiento*, 5. Recuperado de <http://www.um.es/ead/reddusc/5>
- Martín Ciudad, N. (1991). *Factores determinantes de la calidad de la educación universitaria*. En Braza Lloret, P. y otros (Eds.), *Actas del I Congreso Internacional sobre Calidad de la Educación Universitaria*. Cádiz: ICE Universidad de Cádiz.
- Marzo, M., Pedraga, M., y Rivera, P. (2006). Definición y validación de las competencias de los graduados universitarios. *Revista de la Educación Superior*, vol. XXXV, 140, 49-70.
- Montanero, M., et al. (2006). *Orientaciones para la elaboración del Plan Docente de una materia (Guía extensa)*. Oficina de Convergencia Europea. Servicio de Orientación & Formación Docente. Universidad de Extremadura. Recuperado de <http://www.unex.es>
- Mora, J.G. (2003). Competencias y empleo de los jóvenes graduados universitarios. *Revista de Educación*, 330, 157-170.
- Morales, P. (1998). *Evaluación y aprendizajes de calidad*. Guatemala: Universidad Rafael Landívar.
- Muros, B., y Luis, J. C. (2012). Aprendizaje, capacidades cognitivas y evaluación formativa en formación inicial del profesorado. *Revista de Ciencias de la Educación*, 230, 171-182.
- Pérez Juste, R. (2007). La evaluación externa y sus implicaciones. Aspectos técnicos, prácticos y éticos. *Avances en Investigación Educativa*, 6. Recuperado de http://adide.org/revista/index.php?option=com_content&task=view&id=197&Itemid=47
- Rodríguez Ayán, M^a. N., Ruiz, M. Á. (2011). Indicadores de rendimiento de estudiantes universitarios: calificaciones versus créditos acumulados. *Revista de Educación*, 355, 467-492.
- Roca, J. (Coord.) (2008). *Resultados de aprendizaje en las nuevas enseñanzas oficiales de grado*. Proyecto EA 2008-0125, 121-150.

- Sadler, R. (1985). The Origins and Functions of Evaluative Criteria. *Educational Theory*, 35, 285-297. doi:10.1111/j.1741-5446.1985.00285.x
- Sierra, M^a. I. (Coord.) (2008). *Herramientas para la evaluación de competencias en los estudiantes universitarios y recursos para el análisis de su calidad en el marco del Espacio Europeo de Educación Superior*. Proyecto EA2008-0010. Dirección General de Universidades. MEC.
- Smyth, K. (2004). The benefits of students learning about critical evaluation rather than being summatively judged. *Assessment and Evaluation in Higher Education*, 29 (3), 370-378.
- Struyven, K., Dochy, K., y Janssens. S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment and Evaluation in Higher Education*, 30 (4), 325-341.
- Sullivan, K. (2002). Credit and grade transfer within the European's Union SOCRATES Programme: unity in diversity or head in the sand? *Assessment and Evaluation in Higher Education*, 27 (1), 65-74.
- Tabernero, B. y Daniel, M. J. (2012). Análisis de los sistemas de evaluación del alumnado en la formación de maestros: estudio comparativo entre lo que opinan profesores y egresados. *Revista Interuniversitaria de Formación de Profesorado*, 15 (3), 133-144.
- Tejedor, F., García-Valcárcel, A. (2007). Causas del bajo rendimiento del estudiante universitario (en opinión de los profesores y alumnos). Propuestas de mejora en el marco del EEES. *Revista de Educación*, 342, 444-447.
- Tiana, A. (1988): Calidad y evaluación de la educación secundaria, *Comunidad Educativa*, 253, 21-24.
- Tractenberg, R. E., Chaterji, R., y Haramati, A. (2007). Assessing and analyzing change in attitudes in the classroom. *Assessment and Evaluation in Higher Education*, 32(2), 107-120. <http://doi.org/10.1080/02602930600800854>
- Trier, U. y Peschar, J. (1994): *Cross curricular competencias: Rationale and Strategy for Developing a new indicador*. New Orleans, AERA.
- Trotter, E. (2006). Students perceptions of continuous summative assessment. *Assessment and Evaluation in Higher Education*, 31 (5), 505-521.
- Turrull, M., Roca, B., y Alberti, E. (2010). *De nuevo sobre las causas del rendimiento académico. La experiencia de la facultad de derecho de la UB desde la óptica de la gestión académica*. Barcelona: Estudis i anàlisis de la Facultat de Dret.
- Zabalza, M. A., Trillo, F., y Gewerc, A. (Eds.) (2001). La calidad de la docencia en la Universidad. *Actas del I Symposium Iberoamericano sobre Didáctica Universitaria*. Santiago de Compostela.

Weber, L., Frary, R., y Cross, L. (1995). Allowing student a choice of items on objective examination. *Assessment and Evaluation in Higher Education*, 20 (3), 301-306.

Webster, Pepper y Jenkins (2000). Assessing de undergraduate dissertation. *Assessment and Evaluation in Higher Education*, 25 (1), 1-80.

Williams, S. (2006). Assertions-reason multiple-choice testing as a tool for deep learning: a qualitative analysis. *Assessment and Evaluation in Higher Education*, 31 (3), 287-301.