



Artículo de Revisión

La evaluación del aprendizaje universitario: conflictos entre sus distintas utilidades y compromisos de resolución para preservar las características más esenciales en cada tipo de evaluación

Alfredo Prieto ^{1,*}, Isabel Lara ¹, David Díaz ¹, Jorge Monserrat ¹, Ana Pérez-Gómez ¹, Jesús Vélez ², Fernando M. Otero ³

¹ Universidad de Alcalá

² Universidad de Cádiz

³ Universidad Pablo de Olavide, Sevilla

* Autor correspondencia: alfredo.prieto@uah.es

Recibido: 05/04/2018; Aceptado: 26/04/2018; Publicado: 01/05/2018

Resumen: En este artículo examinamos las distintas utilidades de la evaluación del aprendizaje y planteamos los conflictos que puede producir el querer enfatizar demasiado ciertas utilidades secundarias al coste del detrimento de otras más relevantes y características de cada tipo de evaluación. Siguiendo a Cees van der Vleuten (autor premiado en 2012 con el Karolinska Prize for Research in Medical Education) [1] proponemos como solución el llegar a un compromiso fundamentado que nos permita comprometer unas características u otras de la evaluación en función de la utilidad principal que pretendemos lograr por medio de ésta [2,3]. Se revisan las características de los instrumentos de evaluación y se describe la función de utilidad de un método de evaluación. Se discute la dicotomía entre dos tipos de evaluación: la evaluación formativa y de baja repercusión versus la evaluación acreditativa y de alta repercusión. Se establecen las características que son irrenunciables y no se pueden comprometer en cada uno de estos dos tipos de evaluación. Estas características esenciales irrenunciables son la validez, la fiabilidad y la seguridad en la evaluación acreditativa o de alta repercusión (HighStakes Assessment), mientras que en la evaluación formativa de baja repercusión (Low Stakes Assessment) no podemos comprometer el impacto educativo que incluye el estímulo del trabajo del alumno, el ejercicio de competencias y el feedback correctivo.

Palabras Clave: Evaluación del aprendizaje; Evaluación formativa; Evaluación acreditativa; Evaluación de baja repercusión; Evaluación de alta repercusión; Rúbrica; Validez; Fiabilidad; Impacto educativo.

Abstract: In this manuscript the different utilities of learning assessment and conflicts that can be caused by putting excessive emphasis to certain secondary utilities of learning assessment in detriment of another more relevant utilities for a certain type of assessment. Following to assessment expert Cees van der Vleuten [1] (awarded in 2012 with the Karolinska Prize for Research in Medical Education) we propose to reach a compromise which allows us to compromise certain assessment characteristics depending on the main utility of each type of assessment [2, 3]. Characteristics of assessment instruments are reviewed and a function of utility for a given assessment method is described. The dichotomy between the two main types of learning assessment: formative or low stakes assessment versus creditive assessment or high stakes assessment. Essential characteristics which cannot be compromised in each one of those types of assessment are established. These essential characteristics are the validity, the reliability and the security in the summative or High Stakes Assessment, whereas in the formative or Low Stakes

Assessment we cannot compromise the educative impact which includes both the stimulation of student work, the exercising of competences and the administration of corrective feedback.

Key words: Learning assessment; Formative assessment; Accreditive assessment; Low stakes assessment; High stakes assessment; Rubric; Validity; Reliability; Educative impact.

1. Introducción

La evaluación del aprendizaje es un elemento de gran importancia en la enseñanza y la garantía del aprendizaje universitarios. La evaluación estimula esfuerzos de los alumnos que producirán su aprendizaje y a la vez certificará lo que los alumnos han aprendido. La evaluación del aprendizaje es un proceso que sirve múltiples propósitos y esta multifuncionalidad da lugar a conflictos entre las distintas utilidades de la evaluación.

Analizar un método de evaluación en función de su utilidad nos puede ayudar a replantear y mejorar nuestros sistemas de evaluación. Por ejemplo, podemos plantearnos ¿Qué utilidad tiene en el contexto del siglo XXI evaluar la reproducción de la información memorizada y evaluar la reproducción mecánica de procedimientos memorizados para la resolución de problemas tipo? ¿Qué efectos contraproducentes pueden tener estos métodos tradicionales de evaluación en el aprendizaje y en el desarrollo de competencias en los alumnos? ¿Realmente lo que evalúo es una muestra válida y representativa de lo que quiero que aprendan? ¿Ayuda el sistema de evaluación de mis asignaturas a que mis alumnos orienten su concentración y sus esfuerzos hacia los resultados de aprendizaje que sus profesores consideramos prioritarios? ¿Ayuda a que mis alumnos desarrollen su capacidad para la reflexión metacognitiva, la autoevaluación y la autorregulación de su propio aprendizaje? Todas estas son preguntas relevantes si queremos mejorar nuestra evaluación con el fin de mejorar su utilidad y sus efectos positivos sobre el aprendizaje de nuestros alumnos.

El efecto de nuestra evaluación en lo que nuestros alumnos hacen para aprender es de gran importancia. No importa lo innovadora que sea la metodología de enseñanza, lo que determina principalmente cómo trabajan y estudian los alumnos es el método o combinación de métodos con los que vayan a ser evaluados. Es la evaluación la que dirige lo que los estudiantes hacen para superarla, no los buenos deseos o las buenas intenciones de sus profesores, no los objetivos de su asignatura, sino aquellas pruebas y tareas con las que ellos piensan que serán evaluados y calificados [4].

La repercusión de este principio en el aprendizaje de nuestros alumnos es de gran importancia. Si los profesores evaluamos algo distinto de lo que queremos que los alumnos aprendan, en el mejor de los casos, los alumnos se enfocarán en aprender eso que evaluamos y, por tanto, no aprenderán lo que queremos que aprendan. Por ello, es muy importante alinear lo que evaluamos con aquellos resultados de aprendizaje que pretendemos lograr. El alineamiento educativo de la evaluación con los resultados de aprendizaje y actividades de enseñanza aprendizaje implica plantearse preguntas como: ¿Qué es lo que tendría mayor valor comprender y aprender? ¿Está alineado tu sistema de enseñanza y evaluación con los resultados de aprendizaje que pretendes lograr? ¿Qué problemas tenemos para lograr que nuestros alumnos aprendan lo que queremos que aprendan?

Si analizamos la evaluación del aprendizaje desde el punto de vista de su utilidad nos daremos cuenta de que la evaluación del aprendizaje tiene propósitos muy variados, [5]. La evaluación del aprendizaje sirve entre otras cosas para:

- Calificar a los alumnos por medio de la valoración de su aprendizaje y su progreso
- Calificar al profesor y a la asignatura
- Motivar a los alumnos para trabajar
- Proporcionar retroinformación a los alumnos sobre su aprendizaje
- Proporcionar retroinformación al profesor sobre su enseñanza y los problemas de aprendizaje de sus alumnos

- Rendir cuentas sobre la eficacia de la metodología docente
- Mejorar el aprendizaje y la enseñanza

Al preguntar ¿Para qué nos sirve la evaluación? A los profesores participantes en cursos de formación del profesorado universitario sobre evaluación del aprendizaje en España obtenemos respuestas como

- Para calificar a los alumnos (y cumplir con una obligación burocrática)
- Para acreditar, comprobar y medir el aprendizaje que produce la enseñanza
- Para estimular y motivar el esfuerzo de los alumnos
- Para detectar lo que debe corregirse, mejorarse y proporcionar al alumno feedback sobre su aprendizaje
- Para obtener información con la que mejorar el proceso enseñanza aprendizaje y la labor del profesor
- Para proporcionar otra manera de aprender superando retos
- Para entrenar al alumno para que rinda mejor en situaciones de evaluación acreditativa/sumativa que sean similares al procedimiento utilizado para la evaluación formativa.

En ambas listas puede observarse que la evaluación tiene utilidades muy diversas: acreditativas, formativas y de retroinformación al alumno, al profesor y al centro educativo. Aprovechar al máximo todos estos potenciales de la evaluación es clave para mejorar tanto el aprendizaje de nuestros alumnos como el modo de enseñarles. Por tanto, la evaluación del aprendizaje debe servir a un espectro de propósitos que es muy amplio. Estos son tan variados que es muy frecuente que el diseño de la evaluación para el logro de unos propósitos entre en conflicto con el logro de otros propósitos [3]. Al pretender optimizar el logro un propósito diseñamos y hacemos la evaluación de una manera que resulta en efectos contra producentes o evita o dificulta que se logren con ella otros propósitos de gran importancia.

2. La noción de compromiso en la resolución de los conflictos entre las distintas utilidades de la evaluación del aprendizaje

Este conflicto entre las distintas utilidades de la evaluación es de difícil solución y según algunos teóricos la solución reside en la noción del compromiso [3]. La noción de compromiso en general es que no se puede tener todo a la vez, sino que tendremos que conformarnos con tener unos puntos fuertes, siendo conscientes de que esto implica comprometer el logro de otros menos esenciales en esa situación concreta.

Cuando la noción de compromiso se aplica al diseño y la implementación de la evaluación implica que con un determinado tipo o instrumento de evaluación no se puede conseguir optimizar todas las potenciales utilidades de la evaluación. Si queremos maximizar o fortalecer unas utilidades de la evaluación tendremos que sacrificar en otras. En primer lugar, deberemos utilizar y combinar sabiamente distintos instrumentos de evaluación con distintas fortalezas para lograr una combinación óptima de sus utilidades. En segundo lugar, deberemos decidir qué es lo que se puede comprometer y que es lo que no se debe comprometer en cada tipo de evaluación.

Esta sección puede dividirse en sub-apartados. Debe proporcionar una descripción concisa y precisa de los resultados, su interpretación, así como las conclusiones experimentales que se pueden extraer.

2.1. Características de los instrumentos de evaluación

Según van der Vleuten [3], cada instrumento de evaluación tiene una serie de características: validez, fiabilidad, impacto educativo (que incluye el estímulo para el estudio y la oportunidad de

ejercitar competencias generar feedback formativo), aceptabilidad y coste en medios humanos, materiales y económicos

Cees van der Vleuten [3] establece la función de utilidad de un método de evaluación como el resultado de producto de diversos factores. La utilidad de un método de evaluación viene definida por la siguiente función de utilidad:

$$U = p_f F \times p_v V \times p_{ie} IE \times p_a A / p_c C$$

Donde:

U= Utilidad del método o instrumento de evaluación

F = fiabilidad ¿es precisa la medida del aprendizaje? ¿es repetitivo el resultado de la prueba? ¿al volver a examinar al mismo alumno se repite el resultado?

V= validez ¿vale para medir lo que quiero medir o mide otras cosas?

IE = Impacto educativo ¿cómo afecta la aplicación de esta prueba de evaluación al aprendizaje de mis alumnos?

A = Aceptabilidad ¿es aceptable para alumnos y profesores?

C = Coste ¿podemos permitirnos aplicar esta manera de evaluar de forma sostenible con los medios materiales y humanos de los que disponemos?

p= coeficiente de peso relativo que asignamos a cada factor dentro de la función de utilidad

El coeficiente de peso relativo (p) de la fiabilidad (F) y validez (V) debe ser muy alto en la evaluación de utilidad acreditativa/sumativa para que podamos tomar decisiones válidas y fiables sobre la calificación y promoción de nuestros alumnos. Sin embargo, el coeficiente de peso relativo (p) del impacto educativo (IE) debe ser alto en la evaluación de utilidad formativa.

La noción de compromiso en la evaluación significa que no lo podemos tener todo a la vez, debemos por tanto escoger maximizar unas características en detrimento de otras. Para ello, debemos tener muy claro en cada caso cuales son las características esenciales a las que no podemos renunciar en ese tipo de evaluación. En cada tipo de evaluación no podemos sacrificar o comprometer algunas características que son irrenunciables para ese tipo de evaluación. Sin embargo, si podemos comprometer otras que son accesorias e incluso irrelevantes para la función o utilidad principal de ese tipo de evaluación.

3. La dicotomía entre la evaluación formativa y de baja repercusión versus la evaluación acreditativa y de alta repercusión.

Cees van der Vleuten diferencia primero entre la evaluación de utilidad principalmente formativa y aquella cuya función es acreditativa/sumativa y pretende valorar lo aprendido por el estudiante y que servirá de base para la toma de decisiones sobre la calificación del aprendizaje. La solución que nos propone van der Vleuten pasa por diferenciar dos tipos de evaluación del aprendizaje utilizando un sistema de evaluación continua que integra una evaluación formativa de baja repercusión/riesgo (low stakes assessment) con una evaluación acreditativa/sumativa de alta repercusión/riesgo (high stakes assessment) [3].

El primer tipo de evaluación del aprendizaje es la evaluación de baja repercusión (low stakes assessment) cuya utilidad es principalmente formativa (estimuladora de esfuerzo, ejercicio de competencias y ocasión para el feedback formativo), pero que (a diferencia de la evaluación formativa tradicional) también tiene un cierto peso en la calificación (aunque este debe mantenerse en un valor bajo para estimular el esfuerzo, pero no tanto como para propiciar las picardías y comportamientos tramposos por sacar esos puntos). El segundo tipo de evaluación es la evaluación de alta repercusión (high stakes assessment) en la que se pone el énfasis en su validez y fiabilidad para medir el aprendizaje, calificarlo y acreditarlo de manera precisa y exacta.

Esta división de la evaluación en low stakes/high stakes supone un avance conceptual con respecto a la división tradicional entre evaluación formativa (pero de ninguna repercusión) y la acreditativa/sumativa (de total repercusión en la calificación y nulo papel formativo). La división de la evaluación en formativa y acreditativa, suele resultar en que muchos profesores y alumnos no se toman suficientemente en serio la evaluación formativa tradicional, pues por definición no tenía

repercusión en la calificación y, por ello, la mayoría de los alumnos orientaban casi todos sus esfuerzos hacia la superación de la evaluación acreditativa con el mínimo esfuerzo posible. Sin embargo, en el enfoque propuesto basado en distintos niveles de repercusión (low stakes/high stakes), la evaluación formativa tiene algo de repercusión un cierto porcentaje de peso en la calificación (aunque sea bajo), y de este modo será tomada más en serio por los profesores, pero sobre todo por los alumnos que al fin y al cabo son los que deben ser estimulados por la evaluación para trabajar.

4. Aspectos que son esenciales e irrenunciables en cada tipo de evaluación

En la evaluación formativa no podemos comprometer el impacto formativo, esa debe ser la razón por la que realizamos evaluación formativa, queremos que estimule a los alumnos a razonar, a trabajar, a arriesgarse, a errar y a aprender de sus errores por medio del contraste con las respuestas de sus compañeros y el aporte de feedback inmediato. Si además la evaluación formativa es low stakes (y pesa algo en la calificación), los alumnos se la tomarán mucho más en serio que si fuese sólo exclusivamente formativa (zero stakes) y ello redundará en que tendrá aún más impacto sobre el aprendizaje de los alumnos porque les hará trabajar más durante todo el periodo de aprendizaje. Podemos relacionar la idea de low stakes con los sistemas de inducción a la acción mediante pequeñas recompensas, bonificaciones o insignias que se usan en la gamificación para motivar el trabajo continuo de los alumnos [6]. También podemos usar herramientas tecnológicas para la evaluación formativa como las Apps socrative, kahoot, Quizizz o Plickers para gamificar la evaluación formativa en clase con puntuaciones y competiciones divertidas y así convertir la evaluación formativa una actividad entretenida y hasta divertida para nuestros alumnos.

Es muy fácil saber si nuestro sistema de evaluación formativa tiene un impacto deficiente. Si el primer efecto de la evaluación formativa es estimular el trabajo de los alumnos debemos preguntarnos: ¿Logra nuestro sistema de evaluación formativa estimular a nuestros alumnos para que dediquen a nuestra asignatura el volumen de trabajo necesario para alcanzar los resultados de aprendizaje pretendidos? (si/no) ¿Qué progresos obtenemos en el aprendizaje en relación a los niveles de partida hallados en la evaluación diagnóstica? (ganancias de un 20-25% versus ganancias superiores a un 25%) ¿En qué alumnos obtenemos ganancias de aprendizaje más pobres? ¿Qué medios de detección precoz y remedio (individualizado/colectivizado) de alumnos con escaso rendimiento podemos aplicar?

En la evaluación acreditativa/sumativa no podemos comprometer su seguridad, su validez y su fiabilidad. Nuestras pruebas acreditativas deberán ser diseñadas para maximizar su fiabilidad y reproducibilidad (dentro de lo posible). Si no son fiables, no podrán ser válidas y las decisiones acreditativas en base a esas pruebas no fiables dejarán mucho que desear e incurrirán en errores indeseables.

Existe un mito o falsa creencia que considera que algunos métodos de evaluación son "fiables por definición". Sin embargo, para alcanzar la fiabilidad no vale con usar ciertos métodos (como el multiple choice questionnaire, MCQ) sino que es necesario que nuestros sistemas de evaluación (combinaciones de tareas y pruebas) recojan evidencias de una muestra representativa de aquellos conocimientos y habilidades de cuyo aprendizaje se quieren obtener evidencias. Dicho en pocas palabras: la fiabilidad de una evaluación depende principalmente del muestreo, no de la supuesta objetividad del método de calificación. Por ello, usar un método supuestamente objetivo como el examen tipo test MCQ no hace fiable a nuestro método de evaluación si este no realiza un muestreo significativo de aquello que pretendemos evaluar. Incluso métodos considerados más objetivos como el MCQ no son nada fiables a menos que realicen un muestreo representativo. Tampoco serán válidos para evaluar ciertos resultados de aprendizaje a menos que esas preguntas estén asociadas a contextos concretos similares a aquellos en los que los alumnos tengan que transferir su comprensión en la práctica profesional y la vida real.

Un examen de test que sólo muestree durante una hora, simplemente no es fiable ni reproducible por sí sólo para dar una calificación precisa del aprendizaje del alumno. A menos que lo usemos sólo como una evidencia que combinamos con otras fuentes de evidencia para entre todas ellas, construir una estimación del aprendizaje más fiable.

Para que una medida del aprendizaje sea fiable o reproducible debemos estar seguros de que si realizamos otra prueba los resultados serán muy similares a los de la primera. Si el alumno que sacó ayer un 6 saca ahora un 4 al realizar otra prueba similar sobre los mismos contenidos podemos concluir que el resultado del procedimiento de medida no es fiable para decidir si un alumno merece aprobar ¡No nos vale siquiera para tomar decisiones sobre a quién debemos aprobar! Debemos además considerar no sólo el valor máximo de la escala de calificación, sino los mejores resultados obtenidos por nuestros alumnos [7] para decidir si deberíamos ajustar las notas para que el alumno que obtuvo la máxima nota (8,25 por ejemplo) se convierta en un 9 o un 10 y reajustar las notas de sus compañeros a esa nota de referencia máxima.

Hay dos posibles soluciones a este problema de maximizar la fiabilidad. La primera solución que se usa en la universidad y en las pruebas estandarizadas en otros niveles educativos es diseñar la prueba para maximizar su fiabilidad. Se construyen pruebas de alta fiabilidad y máximo poder discriminativo mediante selección de ítems cuya dificultad y poder discriminativo son conocidos por el profesor. De esta manera, podemos construir pruebas de dificultad y poder discriminativo controlados por el profesor.

Para que la evidencia obtenida sea más representativa de lo que el alumno sabe, se muestrean grandes conjuntos de preguntas (hasta 250 en el MIR). Por ello, es conveniente que cuando el objetivo de la evaluación es acreditativo se realicen pruebas de la mayor duración posible (al menos 2 horas para una prueba MCQ) con las que se puede muestrear de manera más fiable. También es importante incluir ítems válidos para evaluar si el alumno sabe aplicar un procedimiento, sabe interpretar y concluir, sabe juzgar, comprende en profundidad y sabe cuándo debe hacerse una determinada intervención.

Sin embargo, si nuestro objetivo es formativo y lo que buscamos es el contraste, la discusión y el feedback formativo lo mejor sería una selección de aquellos ítems que puedan dar lugar a discusiones más vivas, ejemplos que interesen, con probabilidad de acierto del 50% (para generar debate entre los alumnos), y también conjuntos seleccionados de aquellas preguntas más difíciles y discriminativas (por ejemplo las más falladas por los alumnos en exámenes de promociones anteriores) que son las que harán relevante el feedback más necesario sobre aquellos conceptos erróneos más extendidos entre nuestros alumnos.

Por ejemplo, si nuestra intención es formativa podemos poner una prueba de evaluación formativa on line con orden de preguntas barajado para cada alumno y sin feedback automático (para evitar picarescas de copia generalizada) y escoger las preguntas más difíciles (las que más alumnos fallen) para discutir las en clase.

La otra solución para construir una evaluación acreditativa fiable es combinar un número elevado de minipuebas como se hace en los exámenes de evaluación clínica objetiva estructurada (ECO) y a nivel escolar en sistemas de evaluación continua que combinan múltiples fuentes de evidencias de aprendizaje. De esta manera, aunque la fiabilidad de cada prueba individual no sea alta, la fiabilidad de su combinación será mucho mayor y además proporcionará información sobre el proceso de aprendizaje y su avance.

5. Cómo podemos mejorar las distintas utilidades de la evaluación

A continuación, vamos a considerar cómo tenemos que diseñar nuestra evaluación con el fin de fortalecer u optimizar determinados propósitos o utilidades de la misma.

5.1 La evaluación como estímulo y motivación para el que aprende

Cuando lo que pretendemos con nuestras actividades de evaluación es estimular a nuestros alumnos a estudiar y trabajar, es claro que no serán críticas sus propiedades psicométricas de validez y fiabilidad. Entonces podremos comprometer con toda tranquilidad la fiabilidad de las medidas, pero lo que no podremos comprometer será su impacto sobre las actividades de nuestros alumnos. Por ello, no podremos comprometer su efecto estimulante y, por tanto, deberemos escoger maneras de evaluar (y de bonificar a partir de la evaluación) que realmente estimulen al esfuerzo continuado

de nuestros alumnos, aunque no será necesario que esta evaluación nos proporcione medidas precisas y exactas de lo que estos han aprendido.

Los sistemas de bonificación por pequeñas recompensas, por ejemplo, los que se usan para recompensar la preparación para las clases o la participación en ellas asignan pequeños premios en recompensa por acciones que en opinión del profesor conducen al aprendizaje de sus alumnos y al ejercicio de competencias importantes.

Realizar la evaluación formativa de manera divertida también estimulará y motivará al esfuerzo pues cuando los alumnos se divierten deja de operar la ley del mínimo esfuerzo y empieza a aplicar la de la máxima diversión.

Asignar un pequeño peso en la calificación individual a las tutorías de seguimiento de trabajos de equipo estimula que cada miembro del equipo trabaje de manera constante y responsable. De nuevo aquí lo importante será comprobar y recompensar el trabajo de preparación de cada alumno. Sin embargo, no será necesario que se califique la calidad de su trabajo con una gran precisión y exactitud.

Finalmente, es importante que demos a esos resultados de aprendizaje y de desarrollo de competencias que deseamos favorecer, un peso en la calificación de la asignatura que sea acorde y proporcional a su importancia. Así los alumnos priorizarán su esfuerzo y se concentrarán en la consecución de aquellos resultados de aprendizaje más importantes.

5.2 La evaluación como fuente de información y generación de feedback

Cuando el objetivo principal perseguido con la evaluación sea obtener información a tiempo para detectar y resolver las dificultades de nuestros alumnos para comprender y aprender, es claro que deberemos optimizar nuestro método evaluativo para que nos proporcione información relevante sobre cuáles son los conceptos erróneos, las dudas urgentes y las zonas oscuras en la comprensión de nuestros alumnos.

En esta situación la evaluación deberá informarnos (feedforward) de los aspectos claves que todavía no se han aprendido para que el profesor pueda generar retroinformación (feedback) individualizada o colectiva que permitan ayudar a nuestros alumnos a superar estas dificultades.

Los cuestionarios de preguntas reflexivas para la comprobación del estudio previo que se usan en la metodología just-in-time teaching nos ofrecen ese tipo de evaluación anticipada a la interacción presencial que nos suministra información muy útil sobre las reacciones de los alumnos a los materiales de estudio (documentos y vídeos) y permiten que el profesor replantee las clases presenciales después de analizar la información contenida en el feedforward procedente de los alumnos que nos permitirá detectar las dificultades, zonas oscuras y dudas urgentes de sus alumnos tras el estudio previo de los materiales [8,9,10].

Ejemplo de cuestionario de reacción a los materiales para implementar aprendizaje inverso mediante el método *just in time teaching*. [8]

1. Apellidos, Nombre (apellidos primero para que podamos ordenarlos por orden de lista)
2. ¿Cuáles son las ideas que te parecen más interesantes de este tema y justifica por qué?
3. ¿Qué parte necesita profundizarse en clase y justifica por qué?
4. Haz un resumen en 100 palabras de lo más importante del tema
5. ¿Cuál es la parte más oscura o difícil de comprender? ¿Explica por qué?
6. ¿Qué cuestión urgente querrías que te respondiésemos en la primera clase del tema?
7. ¿Qué parte del tema comprendes bien y no necesitas que te expliquen? ¿Por qué?
8. ¿Qué comprendiste al ver los vídeos?
9. Dirección de e-mail (por si te tenemos que pedir que vuelvas a reestudiar los materiales)
10. ¿Sobre qué aspecto o cuestión concreta relacionada con este tema te gustaría realizar una indagación por tu cuenta?

11. Si comprendiste algo importante que no comprendías antes de estudiar los materiales explica que es lo que comprendiste y escribe una cuestión a la que puedan responder los compañeros que también han comprendido esa idea.
12. Haz una breve crítica constructiva de los materiales instructivos proporcionados e indica que cosas podrían ser cambiadas o mejoradas para mejorar la claridad y facilitar la comprensión de las ideas más difíciles de comprender
13. ¿Cuánto tiempo has dedicado a leer, entender y sacar sentido de los documentos y powerpoints facilitados?
14. ¿Cuánto tiempo has dedicado a ver y entender los vídeos?
15. ¿Te fueron útiles las transcripciones de audio de los vídeos?
16. ¿Cuánto tiempo has dedicado a reflexionar y a responder a este cuestionario?

5.3 La evaluación como orientación para el alumno

Tanto John Biggs [4] como Fink [11], Wiggins y Mc Tighe [12, 13] enfatizan la gran importancia de alinear la evaluación con lo que queremos que aprendan nuestros alumnos y con el hecho de comunicarles eficazmente cuáles son esos resultados de aprendizaje deseados y cómo van a ser evidenciados gracias al sistema de evaluación de la asignatura. Por ello, es muy importante transmitir a nuestros alumnos con la máxima transparencia y claridad cuáles serán los resultados de aprendizaje más importantes en nuestra asignatura, con que productos desempeños o pruebas serán evaluados, cuáles serán los criterios usados para evaluar y, finalmente, cuáles serán los indicadores de logro que se utilizarán para valorar el nivel de aprendizaje alcanzado por cada alumno en relación a cada criterio de evaluación. Finalmente, es importante generar actividades de enseñanza-aprendizaje y evaluación formativa en las que los alumnos practiquen las situaciones similares a las que se producirán en la evaluación acreditativa.

5.4 La evaluación formativa como entrenamiento para la evaluación acreditativa y la situación de desempeño real

Sin caer en el extremo de la enseñanza para la prueba y la consiguiente trivialización de la evaluación, es mucho el uso que podemos hacer de la evaluación formativa como ocasión para que los alumnos practiquen y desarrollen habilidades que serán importantes para la superación exitosa de la evaluación acreditativa.

Si queremos que nuestros alumnos resuelvan casos y problemas como profesionales competentes debemos evaluarlos acreditativamente por su capacidad para resolverlos y debemos dedicar tiempo de clase a la práctica de la resolución de este tipo de casos. Debemos introducir en las clases actividades y prácticas que ayuden a nuestros alumnos que aprendan a hacer el tipo de operaciones mentales y prácticas que después evaluaremos en nuestra evaluación acreditativa.

La conclusión es que nuestros alumnos se concentran al máximo y aprenden más cuando comprenden la relación de lo que les proponemos hacer con los resultados de aprendizaje deseados y con los mecanismos de evaluación acreditativa que deberán superar para aprobar la asignatura.

5.5 La evaluación acreditativa/sumativa como instrumento de medida del aprendizaje y de toma de decisiones a partir del nivel de aprendizaje alcanzado

Cuando usamos la evaluación con este propósito son fundamentales las consideraciones acerca de la fiabilidad y la validez. En relación a la validez debemos preguntarnos: ¿Sirve este instrumento para medir el grado de logro del resultado de aprendizaje que pretendo evaluar? El riesgo es que por hacernos más fácil la calificación o por pretender que las pruebas sean más objetivas en muchas ocasiones tendemos a evaluar a niveles de bajo nivel de razonamiento y escasa autenticidad.

Debemos aprender a evaluar en niveles de mayor autenticidad y nivel de razonamiento en los que no tendremos más remedio que hacer juicios subjetivos. Herramientas como las rúbricas nos

ayudarán a explicitar nuestros criterios de evaluación y describir los distintos niveles de ejecución o logro. Las rúbricas también son unas herramientas excelentes para que unos alumnos puedan evaluar el trabajo de sus compañeros y de esta manera pueden aprender a evaluar y colaborar en las tareas de evaluación.

Cuando se pretende medir algo es muy importante saber qué es lo que se pretende medir y si nuestro instrumento es válido para medir precisamente eso que queremos medir. Con el fin de ajustar nuestras herramientas de evaluación a la medición de distintos niveles de razonamiento y de comprensión son muy útiles las taxonomías de Bloom [14, 15] de niveles de razonamiento y de Biggs de estructura de la comprensión [16].

Para saber lo que los alumnos recuerdan debemos preguntarles sobre la información que les hemos transmitido. Para saber lo que los alumnos aplican debemos plantear situaciones de aplicación no mecanizada. Para saber lo que el alumno comprende se usan los tests conceptuales en los que el alumno debe interpretar una situación o contexto, realizar asunciones construir modelos de solución y predecir qué es lo que ocurrirá.

Para saber lo que los alumnos son capaces de hacer, debemos pedirles que hagan o que nos cuenten como lo harían en respuesta a contextos hipotéticos en los que deberían actuar.

Para saber si nuestros alumnos son capaces de transferir su comprensión y esquemas a nuevas situaciones debemos retarles con nuevas situaciones que deberán interpretar y solucionar aplicando aquellos entendimientos que queremos evaluar.

5.6 La evaluación como proceso de aprendizaje y desarrollo de la competencia para la autoevaluación y la autorregulación

Debemos diseñar nuestra evaluación para que obligue a nuestros alumnos a realizar tareas de reflexión metacognitiva y auto evaluación tanto durante el proceso de evaluación formativa como en los exámenes de evaluación acreditativa Nilsson [17]. Construimos así un currículo paralelo en el que nuestros alumnos aprenden a aprender con autodirección y autonomía los contenidos, habilidades y competencias del currículo.

Para que nuestra evaluación logre estos fines. Debes lograr que los alumnos perciban y comprendan el valor intrínseco de las tareas y su relación con los resultados de aprendizaje que se pretenden. Debemos evaluar de otra manera, usando la evaluación compartida [18]. Debemos poner a los alumnos en el trance de tener que evaluar a sus compañeros y autoevaluarse (a esto nos ayudará el proporcionarles rúbricas en las que visualizarán los criterios y niveles de desempeño con los que podrán juzgar el desempeño de sus compañeros). Debemos usar exámenes y quizzes formativamente para producir aprendizaje y crear oportunidades para la discusión con los compañeros y el feedback. Finalmente, debemos usar tareas de reflexión metacognitiva y portafolios para que tus alumnos tengan que reflexionar sobre lo que aprenden y sobre cómo pueden demostrarlo.

5.7 La evaluación como proceso que informa sobre cómo mejorar la enseñanza

En su sentido más amplio la evaluación formativa tiene como finalidad mejorar los procesos de enseñanza aprendizaje al lograr que 1) el alumno aprenda más, 2) que el profesorado mejore su práctica docente y 3) poder mejorar sobre la marcha el proceso de enseñanza-aprendizaje [18].

A este fin contribuye el análisis crítico de la evaluación sobre el aprendizaje. Si analizamos los resultados de la evaluación del aprendizaje no desde la perspectiva de calificar a cada alumno sino desde la de analizar el nivel de consecución en distintos resultados de aprendizaje por una clase durante o tras el periodo formativo podremos tener una medida del nivel de éxito en la consecución de cada uno de los resultados de aprendizaje que se pretendían lograr. Existe una metodología de determinación de ganancias de aprendizaje que permite conocer y comparar el nivel de progreso logrado en la consecución de distintos resultados de aprendizaje. En esta metodología se realiza una evaluación diagnóstica inicial que se repite por sorpresa al final del periodo formativo. De esta manera podemos determinar la ganancia de aprendizaje en cada alumno y la ganancia media en el grupo de alumnos.

Otra manera de evaluar para mejorar la enseñanza es pedir a los alumnos que nos informen directamente sobre lo que comprenden y lo que no. Que nos comuniquen sus dudas acuciantes y tras la experiencia formativa nos aporten sugerencias sobre como la experiencia de aprendizaje podría ser mejorada. Ver nuestra enseñanza desde la perspectiva de nuestros alumnos nos ayudará a detectar problemas que nos pasaban inadvertidos y nos pone en el camino de solucionarlos [19] (Harris Haydn 2006).

6. Conclusiones

La evaluación del aprendizaje es un proceso de gran importancia para que éste se produzca. Las utilidades de la evaluación del aprendizaje son tan diversas que con frecuencia se producen conflictos de intereses en el logro de las características más deseables sobre el aprendizaje de los alumnos. Por ejemplo, al pretender mantener altos niveles de la fiabilidad calificativa en la evaluación formativa, aumentamos su coste en medios humanos con lo que nos vemos obligados a reducir su frecuencia y con ello comprometemos también la frecuencia del suministro de feedback con lo que reducimos su impacto educativo, que es la característica más importante de este tipo de evaluación. Otros problemas se producen en la evaluación acreditativa, que en ocasiones dividimos para cumplir con las normativas de los grados del EEES. Esta división, si se hace sin cuidado, provoca que la evaluación acreditativa se realice en condiciones de fiabilidad y seguridad insuficientes para cumplir su propósito acreditativo de juzgar con validez y fiabilidad el aprendizaje que nuestros alumnos han alcanzado.

Para mejorar este estado de cosas es necesario que los profesores comprendan las características de los instrumentos de evaluación: validez (autenticidad), fiabilidad (reproducibilidad y seguridad), impacto educativo (estímulo del trabajo del alumno, ejercicio de competencias y ocasión para la generación de feedback correctivo), aceptabilidad y coste en medios humanos y materiales. Es necesario adoptar una solución de compromiso en la que cada tipo de evaluación tiene unas características esenciales que no se pueden comprometer. Estas características esenciales irrenunciables son la validez, la fiabilidad y la seguridad en la evaluación acreditativa o de alta repercusión (HSA), mientras que en la evaluación formativa de baja repercusión (LSA) no podemos comprometer el impacto educativo que incluye el estímulo del trabajo del alumno, el ejercicio de competencias y el feedback correctivo.

Agradecimientos: A la Universitat de Lleida por promover y mantener a lo largo de la última década cursos de formación del profesorado sobre evaluación de aprendizaje que impulsaron a los autores de este artículo a revisar durante años sucesivos este campo de estudio. A cees van der Vleuten por sus extraordinarias aportaciones al campo de la evaluación del aprendizaje universitario en Medicina

Conflictos de Intereses: los autores no declaran conflicto de intereses.

Abreviaturas

Las siguientes abreviaturas son usadas en este manuscrito:

LSA Low stakes assessment.

HAS High stakes assessment.

MCQ multiple choice questionnaire.

Referencias Bibliográficas

- 1 Cees van der Vleuten is awarded Karolinska Institutet Prize for Research in Medical Education Karolinska Institutet News [Internet]. 2012 Disponible en <https://ki.se/en/news/cees-van-der-vleuten-is-awarded-karolinska-institutet-prize-for-research-in-medical-education> [Consultado el 4 de abril 2018]
- 2 Van der Vleuten CPM The assessment of professional competence: developments, research and practical implications, 1996 *Advances in Health Sciences Education* 1 (1), 41-67.
- 3 Assessment of the Future, Cees van der Vleuten Karolinska Institutet PRIME Symposium [Internet].2012 <https://www.youtube.com/watch?v=bvFbmTRVjIE&t=1332s> [Consultado el 4 de abril 2018]
- 4 Biggs, J.. Calidad del aprendizaje universitario. 2006 Narcea Madrid
- 5 Mazur, E assessment the silent killer of learning [Internet]. 2013. Dudley Herschbach Teacher/Scientist Lecture <https://www.youtube.com/watch?v=CBzn9RAJG6Q> [Consultado el 4 de abril 2018]
- 6 Chou, Y. Actionable Gamification: Beyond Points, Badges, and Leaderboards 2015 Kindle Edition Octalysis Media.
- 7 Cohen-Schotanus J & van der Vleuten CPM A standard setting method with the best performing students as point of reference: Practical and affordable 2010 *Med Teach.*;32(2):154-60
- 8 Prieto A, Barbarroja J, Cano I Díaz D Lara I Monserrat J Sanvicén P y Vélez J. Flipped learning: aplicar el modelo de aprendizaje inverso 2017 Narcea Ediciones. Madrid
- 9 Prieto A., Díaz D., Monserrat J., Reyes E. Experiencias de aplicación de estrategias de gamificación a entornos de aprendizaje universitario. 2014 *Revisión*, 7(2), 76-92.
- 10 Nuevas combinaciones de aula inversa con just in time teaching y análisis de respuestas de los alumnos. Prieto A, Díaz D, Lara I, Monserrat J, Sanvicen P, Santiago R, Corell A, Álvarez-Mon M. RIED. Revista Iberoamericana de Educación a Distancia, [S.l.], v. 21, n. 1, p. 175-194, ene. 2018. ISSN 1390-3306. Disponible en: <<http://revistas.uned.es/index.php/ried/article/view/18836/16908>>. Fecha de acceso: 04 abr. 2018 doi:<https://doi.org/10.5944/ried.21.1.18836>.
- 11 Fink LD Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses 2nd Edition 2013 John Wiley and Sons San Francisco CA
- 12 Wiggins G y Mc Tighe J The Understanding by Design Guide to Creating High-Quality Units. 2005 ASCD Alexandria
- 13 McTighe, J., & Wiggins, G. Essential questions: opening doors to student understanding. 2013 ASCD Alexandria.
- 14 Bloom, B.S., Englehart, M.D., Furst, E.J., & Krathwohl, D.R. Taxonomy of educational objectives: Cognitive domain. 1956 New York: McKay.
- 15 Churches, A. Edorigami, Blooms taxonomy and digital approaches [Internet]. 2007. <http://edorigami.wikispaces.com/Bloom%27s+and+ICT+tools> [Consultado el 4 de abril 2018]
- 16 Biggs, J.B., and Collis, K.F. Evaluating the Quality of Learning-the SOLO Taxonomy (1st ed) 1982 New York: Academic Press.
- 17 Nilsson L Creating self regulated learners 2013 Stylus Publishing Sterling Virginia
- 18 López-Pastor V y Pérez-Pueyo Evaluación formativa y compartida en educación. 2017 Servicio de Publicaciones de la Universidad de León
- 19 Harris R & Haydn T Pupils' enjoyment of history: what lessons can teachers learn from their pupils? 2006 *The curriculum Journal* 17(4): 315-333.



© 2018 por los autores; Esta obra está sujeta a la licencia de Reconocimiento 4.0 Internacional de Creative Commons. Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by-nc-nd/4.0/>.